

IRE Transactions



on ANTENNAS and PROPAGATION

Volume AP-7

DECEMBER, 1959

Special Supplement

SYMPOSIUM
ON
ELECTROMAGNETIC THEORY

Report of Conference
held
June 15-20, 1959
at the
University of Toronto

Sponsored by
Commission VI of URSI
and the
University of Toronto

PUBLISHED BY THE
Professional Group on Antennas and Propagation

IRE PROFESSIONAL GROUP ON ANTENNAS AND PROPAGATION

Administrative Committee

Arthur Dorne, *Chairman*

E. C. Jordan, *Vice-Chairman*

K. S. Kelleher, *Secretary*

S. Bowhill

R. C. Hansen

W. H. Radford

R. N. Bracewell

S. M. King

E. K. Smith

J. W. Findlay

R. K. Moore

K. M. Siegel

H. Fine

O. G. Villard, Jr.

Ex-Officio Members

J. I. Bohnert

Arthur Dorne

R. L. Mattingly

H. G. Booker

F. T. Haddock

D. C. Ports

J. W. Herbstreit

Honorary Member

L. C. Van Atta

Chapter Chairmen

Albuquerque-Los Alamos

J. R. Ames

Dayton

C. G. Conrad

San Diego

B. I. Small

Akron

H. F. Mathis

Denver-Boulder

J. W. Herbstreit

San Francisco

N. J. Gamara

Boston

J. Walsh

Los Angeles

L. A. Kurtz

Syracuse

D. K. Cheng

Chicago

H. L. Woodbury

Orange Belt

J. D. Montgomery, Jr.

Washington, D. C.

H. T. Ward

Philadelphia

D. R. Crosby

J. B. Smyth, *Editor*

S. A. Bowhill, *Chairman, Papers Review Committee (Propagation)*

H. V. Cottony, *Chairman, Papers Review Committee (Antennas)*

Copies of this special supplement can be purchased from THE INSTITUTE OF RADIO ENGINEERS, 1 East 79 St., New York 21, N.Y. PRICE PER COPY: members of the Professional Group on Antennas and Propagation, \$8.00; members of the IRE, \$12.00; libraries, \$12.00; nonmembers, \$16.00. IRE TRANSACTIONS ON ANTENNAS AND PROPAGATION. Copyright 1960, by The Institute of Radio Engineers, Inc. Printed in U.S.A.

**PROCEEDINGS OF
THE SYMPOSIUM ON ELECTROMAGNETIC THEORY**

THE UNIVERSITY OF TORONTO

Toronto, Ontario, Canada

June 15-20, 1959

**Published as a Special Supplement to the IRE TRANSACTIONS ON
ANTENNAS AND PROPAGATION, Volume AP-7 (1959).**

SYMPOSIUM COMMITTEES

Organizing Committee

GEORGE SINCLAIR, *Chairman*, University of Toronto, Toronto, Canada
S. SILVER, University of California, Berkeley, California
K. M. SIEGEL, University of Michigan, Ann Arbor, Michigan
N. MARCUVITZ, Polytechnic Institute of Brooklyn, Brooklyn, New York
J. R. WAIT, National Bureau of Standards, Boulder, Colorado
E. C. JORDAN, University of Illinois, Urbana, Illinois
R. L. MATTINGLY, Bell Telephone Laboratories, Inc., Whippany, New Jersey

Arrangements Committee

GEORGE SINCLAIR, *Chairman*
VICTOR G. SMITH, *Vice-Chairman and Treasurer*
J. E. REID, *Social Activities*
G. R. SLEMON, *Publications*
J. L. YEN, *Facilities*
A. J. KRAVETZ, *Accommodations*
P. BIRINGER, *Registration*
H. COURTICE, *Transportation and Preprints*
G. F. TRACY, *General Arrangements*

Editorial Board

E. C. JORDAN, *Editor*
MORRIS KLINE
N. MARCUVITZ
G. A. MILLER

K. M. SIEGEL
G. SINCLAIR
J. R. WAIT
J. B. SMYTH (*Ex Officio*)

TABLE OF CONTENTS

| | | |
|------------------------|------------------|----|
| Welcoming Address..... | <i>S. Silver</i> | S5 |
|------------------------|------------------|----|

DIFFRACTION AND SCATTERING THEORY

| | | |
|---|--|------|
| Infinite Integral Transforms in Diffraction Theory..... | <i>P. C. Clemmow</i> | S7 |
| Diffraction of Scalar Waves by a Circular Aperture..... | <i>J. Bazer and A. Brown</i> | S12 |
| Scalar Diffraction by an Elliptic Cylinder..... | <i>N. D. Kazarinoff and R. K. Ritt</i> | S21 |
| Fock Theory—An Appraisal and Exposition..... | <i>R. F. Goodrich</i> | S28 |
| Reduction of the Integral Equations for High-Frequency Diffraction by Disks and Strips..... | <i>B. Noble</i> | S37 |
| Pulse Return from a Sphere..... | <i>V. H. Weston</i> | S43 |
| Decay Exponents and Diffraction Coefficients for Surface Waves on Surfaces of Nonconstant Curvature..... | <i>J. B. Keller and B. R. Levy</i> | S52 |
| New Results in Backscattering from Cones and Spheroids..... | <i>A. Olte and S. Silver</i> | S61 |
| Diffraction by Surfaces of Variable Curvature..... | <i>W. Franz and K. Klante</i> | S68 |
| Diffraction of an Electromagnetic Plane Wave by a Funnel-Shaped Screen..... | <i>Werner Braunbek</i> | S71 |
| The Experimental Determination of the Far-Field Scattering from Simple Shapes (<i>Abstract</i>)..... | <i>J. E. Keys and R. I. Primich</i> | S77 |
| The Diffraction and Refraction of Plane Pulses..... | <i>V. M. Papadopoulos</i> | S78 |
| The Field of a Pulsed Dipole in an Interface (<i>Abstract</i>)..... | <i>C. S. Gardner and J. B. Keller</i> | S87 |
| Diffraction by a Half-Plane with a Special Impedance Variation..... | <i>J. Shmoy</i> | S88 |
| A New Method for the Determination of Far Fields with Applications to the Problem of Radiation of a Line Source at the Tip of an Absorbing Wedge..... | <i>S. N. Karp and F. C. Karal, Jr.</i> | S91 |
| Some New Forms of Huygens' Principle..... | <i>V. H. Rumsey</i> | S103 |
| A Solution to the Equiangular Spiral Antenna Problem (<i>Abstract</i>)..... | <i>V. H. Rumsey</i> | S117 |
| General Theorems on the Transmission Coefficient from a Transmitting to a Receiving System..... | <i>Jean Robieux</i> | S118 |
| On Helmholtz's Theorem in Finite Regions (<i>Digest of Paper</i>)..... | <i>J. Van Bladel</i> | S119 |

RADIO TELESCOPES

| | | |
|--|---|------|
| The Synthesis of Large Radio Telescopes by the Use of Radio Interferometers..... | <i>M. Ryle, A. Hewish, and J. R. Shakeshaft</i> | S120 |
| Experimental Test of a Stepped Zone Mirror for Microwaves..... | <i>G. Toraldo di Francia, L. Ronchi, and V. Russo</i> | S125 |

SURFACE WAVES

| | | |
|--|---|------|
| Preface to the Surface Wave Papers..... | <i>James R. Wait</i> | S132 |
| Anatomy of "Surface Waves"..... | <i>S. A. Schelkunoff</i> | S133 |
| Waves on Interfaces..... | <i>Georg Goubau</i> | S140 |
| Surface Waves Supported by Cylindrical Surfaces..... | <i>H. M. Barlow</i> | S147 |
| Guiding of Electromagnetic Waves by Uniformly Rough Surfaces..... | <i>James R. Wait</i> | S154 |
| Some Theoretical Results for Surface Wave Launchers..... | <i>J. Brown</i> | S169 |
| The Surface-Wave Concept in Connection with Propagation Trajectories Associated with the Sommerfeld Problem..... | <i>H. Bremmer</i> | S175 |
| The Transmission Characteristics of a Corrugated Guide..... | <i>Gerhard Piefke</i> | S183 |
| Radiation and Guided Waves..... | <i>A. E. Karbowiak</i> | S191 |
| Guided Waves on Sinusoidally-Modulated Reactance Surfaces..... | <i>A. A. Oliner and A. Hessel</i> | S201 |
| On the Excitation of the Waves of Proper Solutions..... | <i>Koichi Furutsu</i> | S209 |
| Surface-Wave Research in Sheffield..... | <i>M. F. Bracey, A. L. Cullen, E. F. F. Gillespie, and J. A. Staniforth</i> | S219 |
| Diffraction by Smooth Conical Obstacles (<i>Abstract</i>)..... | <i>H. E. J. Neugebauer and M. P. Bachynski</i> | S226 |
| Surface Waves Over a Lossy Conductor..... | <i>Bernard Friedman</i> | S227 |
| Electromagnetic Properties of Wedge and Cone Surfaces with a Linearly Varying Surface Impedance..... | <i>L. B. Felsen</i> | S231 |

BOUNDARY VALUE PROBLEMS

| | | |
|--|--|------|
| The Finite Range Wiener-Hopf Integral Equation and a Boundary Value Problem in a Waveguide..... | <i>Raj Mittra</i> | S244 |
| Fields in the Neighborhood of a Caustic..... | <i>Irvin Kay</i> | S255 |
| On the Discontinuity Problem at the Input to an Anisotropic Waveguide..... | <i>A. D. Bresler</i> | S261 |
| A Ferrite-Filled Cylindrical Cavity..... | <i>T. S. Chu and R. G. Kouyoumjian</i> | S273 |
| Attenuation in Wedge and Septate Waveguides..... | <i>Robin N. Chisholm</i> | S279 |
| Diffraction of Nearly Plane 3.2-cm EM Waves by 45° and 90° Conducting Wedges—Comparison with Theory..... | <i>N. E. Hedgecock and A. B. McLay</i> | S284 |
| The Matching of Parallel Dielectric Plates to Free Space..... | <i>G. C. McCormick</i> | S288 |

THE PROPAGATION OF WAVES THROUGH VARIOUS MEDIA

| | | |
|---|--|------|
| On the Propagation of Electromagnetic Waves Through Anisotropic Layers..... | <i>G. Tyras and G. Held</i> | S296 |
| The Electromagnetic Field in a Randomly Inhomogeneous Medium..... | <i>W. C. Hoffman</i> | S301 |
| Scattering by Quasi-Periodic and Quasi-Random Distributions..... | <i>V. Twersky</i> | S307 |
| Modified WKB Methods for the Propagation and Scattering of Electromagnetic Waves..... | <i>D. S. Saxon</i> | S320 |
| Interaction of Electromagnetic Waves with Some Natural Surfaces..... | <i>William H. Peake</i> | S324 |
| Electromagnetic Scattering by High-Density Meteor Trails..... | <i>H. Brysk</i> | S330 |
| Electromagnetic Properties of High-Temperature Air..... | <i>M. P. Bachynski, T. W. Johnston, and I. P. Shkarofsky</i> | S337 |
| The Propagation of Electromagnetic Waves in Ionized Gases..... | <i>F. H. Northover</i> | S340 |

ANTENNAS

| | | |
|---|---|------|
| Antennas on Circular Cylinders..... | <i>H. Loltrup Knudsen</i> | S361 |
| Impedance Properties of Complementary Multiterminal Planar Structures..... | <i>Georges A. Deschamps</i> | S371 |
| The Bandwidth of Helical Antennas..... | <i>T. S. M. Maclean and R. G. Kouyoumjian</i> | S379 |
| Numerical Integration Methods for Antenna Pattern Calculations..... | <i>Charles C. Allen</i> | S387 |
| The Numerical Solution of Antenna and Scattering Problems..... | <i>George Sinclair</i> | S402 |
| The Finite Conical Antenna..... | <i>S. Adachi, R. G. Kouyoumjian, and R. G. Van Sickle</i> | S406 |
| Broad-Band Multislot Antenna..... | <i>P. Marie</i> | S412 |
| Resonance and Supergain Effects in Small Ferromagnetically or Dielectrically Loaded Biconical Antennas..... | <i>Charles Polk</i> | S414 |
| The Calculated Phase Velocity of Long End-Fire Uniform Dipole Arrays..... | <i>F. Serracchioli and C. A. Levis</i> | S424 |
| Network Theory and Its Relation to the Theory of Linear Systems..... | <i>J. Meixner</i> | S435 |
| Linear Arrays: Currents, Impedances, and Fields, I..... | <i>Ronald King</i> | S440 |
| A New Method of Near Field Analysis..... | <i>R. C. Hansen and L. L. Bailin</i> | S458 |
| Back Scattering at High Frequencies from a Conducting Cylinder with Dielectric Sleeve..... | <i>Ralph D. Kodis</i> | S468 |
| Modes in Rectangular Guides Partially Filled with Transversely Magnetized Ferrite..... | <i>G. Barzilai and G. Gerosa</i> | S471 |
| Asymptotically Expansible Solutions of the Helmholtz Equation (<i>Abstract</i>)..... | <i>Krystyn Bochenek</i> | S475 |
| Index..... | Follows page | S475 |

Welcoming Address

S. SILVER†

IT is my pleasure to welcome you on behalf of URSI to the Electromagnetic Wave Theory Symposium.

This is, as you know, the third in the series of symposia on this topic initiated by Commission VI of the International Scientific Radio Union. The concept of these symposia was born at the 1952 General Assembly of URSI in Sydney, Australia, and was given its first material existence the following year at McGill University, Montreal, Canada. With phenomenal growth, it attained its stature at the 1955 meeting at the University of Michigan in Ann Arbor. It is certainly fitting that now, in this phase of youthful maturity, it should again be held in Canada before starting (as we hope it will) on its tour through the some thirty nations who are currently members of the Union. It is with a sense of indebtedness and gratitude that I express on behalf of Commission VI thanks and appreciation to the University of Toronto for extending its hospitality to us, and to Prof. Sinclair for undertaking the organization of this symposium. I am also pleased to express appreciation to the IRE Professional Group on Antennas and Propagation for its sponsorship of the Symposium, and to the National Research Council of Canada and to the U. S. National Science Foundation, whose financial support make this meeting possible.

On an occasion such as this, one is tempted to relate the work of the present to that which has gone before and to comment on the contributions of Maxwell and other giants of the past. However, in this symposium, I am sure that Maxwell will not suffer by going unmentioned (as he does in the indexes of more than one well-known work in electromagnetic theory); and, if I can extrapolate from the experience of previous meetings, I am sure that at least one of us will receive his comeuppance during the week by being informed that "Raleigh already did it." I shall, therefore, forbear making a historical review of the field, and comment briefly on the immediate purpose and direction of our program.

The subjects chosen for the symposium stem from the discussions held at the last General Assembly in Boulder, Colo. The central problem in the theory of scattering and diffraction in the form of a boundary value problem remains that of finding a methodology for closing the gap between the low-frequency and high-frequency asymptotic representations of solutions. In fact, there even remains the matter of obtaining the high-frequency representation in a direct manner from

the general formulation of the solution, rather than by the semi-heuristic methods now being employed for bodies other than the cylinder and the sphere. We hope that the results presented at this meeting will show some progress toward resolving this very difficult problem.

The subject of surface waves loomed large in the Boulder meeting. It is perhaps amusing that after all the years of intensive work that have gone into the field, there are still differences of opinion as to the existence or nonexistence of surface waves under certain conditions. The difference in opinion may be largely one of semantics, but it does lead to serious difficulties. However, of a more serious nature is the problem of assuring that in a given situation, one does indeed have the complete solution, and has not by some process of partitioning and of mathematical juggling over-emphasized an artificial aspect of the solution. A study group was formed at Boulder with Dr. Wait as chairman for the purpose of reviewing the field, and part of the program of this meeting is an outgrowth of that activity.

The solutions of the problems of scattering by isolated objects, aside from their intrinsic interest, are also necessary steps in treating scattering by many bodies. For many-body situations, we still have with us the open scattering problems associated with the individual objects that make up the collection, plus the new problems arising from their fixed or statistical spatial distribution. Much has been done in recent years to develop self-consistent general treatments of periodic and "gas-like" distributions, and of arbitrary configurations of widely separated objects, which hold independently of the component scatterers; in these developments, the spatial distribution is treated explicitly for once and for all, and the results for the isolated scatterers (which are essentially "parameters" of the problem) can be inserted for particular applications. Most recently, multiple scattering procedures have been extended to densely packed cases and to "quasi-random" and "quasi-periodic" distributions, and we shall hear about some of the new results at this meeting.

The single and many-body scattering problems reach into various fields of radio, from propagation through the atmosphere to the use of satellites as communication devices. This brings us to an important point that bears on the future course of Commission VI: at the Washington meeting of the USA National Committee this past spring, Dr. Berkner, president of URSI, delivered an address on the impact of space research and space communications on the future programs of URSI. It is easy to see that whereas in the past, Commission VI had a sort of secondary place in the over-all program of

† Chairman of International Commission VI of URSI; Director, Electronics Res. Lab., Dept. of Elec. Engrg., University of California, Berkeley, Calif.

URSI, it will now assume a far more important position. For example, in the area of space communications, whether we consider systems of satellites serving as passive scatterers or systems of relay stations, the problems which we are treating in this symposium become of great importance. The relevance of solutions of scattering problems to the satellite system is quite self-evident. Also, however, we should realize that we must take another look at antenna problems in general. The bandwidth of the space communication channel will be no greater than the bandwidth of the antennas involved. The problems range from those involved in communication between two points on the earth via a satellite system, to communication between a point in the earth and a space vehicle.

The degradation of intelligence arising from the limitations of the antenna characteristics or from the spectral dispersion of the signal caused by scattering proc-

esses must be studied, and new techniques in which the antenna is approached from a data-processing point of view must be considered. It appears to me that the electromagnetic wave theory problems which we shall be treating in this symposium are just the beginning steps to be taken in the solution of larger problems. We were, of course, aware of those larger problems previously; but then they were essentially academic ones, and only now, under the impact of the developments of space technology of the past two years, have they acquired practical significance.

I have touched upon the space communication and space relay subject only to point up the importance of this symposium to the work of Commission VI in the future. The program of this meeting is itself an expression of the vitality of the field and it appears that this is just the beginning. I know that we shall have an interesting and fruitful week.

Infinite Integral Transforms in Diffraction Theory*

P. C. CLEMMOW†

Summary—It is emphasized that rather flexible general solutions of the scalar wave equation formed by integrating a separable solution over a separation "constant" are of direct application in certain boundary value problems if an appropriate infinite integral transform is available. The theory, on these lines, of the diffraction of a scalar plane wave in turn by a circular cylinder, a sphere, and a parabolic cylinder is briefly indicated.

I. INTRODUCTION

PROBABLY most attempts to solve boundary problems start exactly with what appears *a priori* to be suitable general solution of the differential equation and then, if possible, use the restrictions imposed by the boundary conditions to pick out the particular solution. The classical situation is that in which the boundaries are complete coordinate surfaces in a separable coordinate system. The classical method, in this case, is to take the general solution to be an infinite series in which the terms, with arbitrary coefficients, are separable solutions and the summation is over integral values of a separation constant; the coefficients are then determined from the boundary conditions.

In diffraction theory it has long been recognized, explicitly from circular cylinder and sphere problems, that the classical solution, though readily obtained, is only of value as it stands when the diffracting obstacle is not large compared to the wavelength; for the number of terms of the series required to give a good approximation to the infinite sum is of the order of 2π times the ratio of the linear dimensions of the obstacle to the wavelength. It is now appreciated that a more flexible general solution is an infinite integral, in which the integrand is a separable solution multiplied by an arbitrary function of a separation "constant," the latter being the variable of integration. The purpose of the present paper is to indicate, by examples, that a general solution of this form can readily be applied to diffraction problems provided that the appropriate infinite integral transforms are available.

The scalar diffraction of a plane wave in turn by a circular cylinder, a sphere, and a parabolic cylinder is briefly treated below. The respective classical solutions to these problems involve expansions in terms of harmonic functions, of Legendre polynomials, and of Hermite polynomials, the coefficients of which expansions can be regarded as *finite* integral transforms. Here, solutions are obtained by using the inverses of the corresponding *infinite* integral transforms. The relevant Fourier analysis is, of course, well known; the Legendre analysis is

thought to be new,¹ though similar in some respects to that used by Felsen;² and that associated with the parabolic cylinder function is due to Erdelyi and Cherry.³

No details of the behavior of the solutions are given here. For the circular cylinder and sphere, those pertaining to the present method are set out elsewhere^{4,5} and others have previously been discussed in many papers by authors using different methods. For the parabolic cylinder, some are to be found in a rather comprehensive paper by Rice.⁶ Rice's starting point is an application of what may be called the Watson technique to the classical series solution, but he also indicates a way in which his basic result could be obtained more quickly by an appeal, in essence, to the Erdelyi-Cherry transform. Even this latter derivation, however, is puzzling in its reliance on a trick, whereas the derivation given in the present paper is quite straightforward.

It should be stressed that there are, of course, other well-known ways of setting up solutions to diffraction problems which achieve essentially the same ends. Which is "best" seems to the author to be, at least in part, a subjective matter, and no serious attempt at comparison is made here. It is perhaps helpful, however, to remark that the present method has some affinity with that advocated by Felsen,^{2,7} particularly with regard to the use, where appropriate, of an angular variable whose values extend into "nonphysical" space; and that, specifically for the circular cylinder, there is a hint of the same approach in an early paper by Debye,⁸ and a close parallel in Friedlander's work on pulses.⁹

II. GENERAL PRELIMINARY REMARKS

In this section some prefatory remarks are offered, which, although rather elementary and covering well-

¹ P. C. Clemmow, "Studies in Radar Cross Sections XXXIV—An Infinite Legendre Integral Transform and Its Inverse," Rad. Lab., University of Michigan, Ann Arbor, Mich., Rept. No. 2778-5-T; March, 1959.

² L. B. Felsen, "Radiation from ring sources in the presence of a semi-infinite cone," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-7, pp. 168–180; April, 1959.

³ T. M. Cherry, "Expansions in terms of parabolic cylinder functions," *Proc. Edinburgh Math. Soc.*, vol. 8, pp. 50–65; 1949.

⁴ P. C. Clemmow, "Studies in Radar Cross Sections XXXII—On the Theory of the Diffraction of a Plane Wave by a Large Perfectly Conducting Circular Cylinder," Rad. Lab., University of Michigan, Ann Arbor, Mich., Rept. No. 2778-3-T; February, 1959.

⁵ P. C. Clemmow, "Studies in Radar Cross Sections XXXV—On the Scalar Theory of the Diffraction of a Plane Wave by a Large Sphere," Rad. Lab., University of Michigan, Ann Arbor, Mich., Report No. 2778-6-T; April, 1959.

⁶ S. O. Rice, "Diffraction of plane radio waves by a parabolic cylinder," *Bell. Sys. Tech. J.*, vol. 33, pp. 417–504; 1954.

⁷ L. B. Felsen, "Alternative field representations in regions bounded by spheres, cones, and planes," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-5, pp. 109–121; January, 1957.

⁸ P. Debye, "Das elektromagnetische Feld um einen Zylinder und die Theorie des Regenbogens," *Physik. Z.*, vol. 9, pp. 775–779; 1908.

⁹ F. G. Friedlander, "Diffraction of pulses by a circular cylinder," *Commun. in Pure and Appl. Math.*, vol. 7, pp. 705–732; 1954.

* The research reported in this paper has been sponsored by the Rome Air Dev. Center, Air Res. and Dev. Command, under Contract AF 30(602)-1853.

† Radiation Laboratory, University of Michigan, Ann Arbor, Mich.

trodden ground, are nevertheless useful for later reference. First, the classical solution of the circular cylinder problem is derived in a way particularly suited to comparison with the treatment of the same problem given in Section III. Then, some features of the relation between Fourier integral transforms and solutions of the wave equation in rectangular Cartesian coordinates are emphasized, again as a guide to the subsequent analysis.

Throughout this section, the time factor $\exp(i\omega t)$ is suppressed, and for the sake of definiteness, only the scalar wave equation

$$\nabla^2 U + k^2 U = 0, \quad (1)$$

with boundary condition $U=0$ is discussed, where ω and k are real positive numbers.

Consider, then, the plane wave

$$U^i = e^{-ikr \cos \theta} \quad (2)$$

incident on the circular cylinder $r=a$, where r and θ ($0 \leq \theta < 2\pi$) are two-dimensional cylindrical polar coordinates. The solution of (1) separable in r and θ which is adopted in the classical analysis is

$$H_n^{(2)}(kr)e^{in\theta}. \quad (3)$$

Here the Hankel function of the second kind insures that the solution has the outgoing radiation character appropriate to the scattered field; and n is taken to be an integer, or zero, on the grounds that the field is a periodic function of θ , of period 2π .

The scattered field is written in the general form

$$U^s = - \sum_{n=-\infty}^{\infty} \frac{a_n}{H_n^{(2)}(ka)} H_n^{(2)}(kr)e^{in\theta}, \quad (4)$$

where the coefficients a_n are to be determined from the boundary condition, and the factor $-1/H_n^{(2)}(ka)$ is inserted in (4) merely to make this assume the most convenient form. In fact, equating to zero the sum of (2) and (4) on $r=a$ gives

$$\sum_{n=-\infty}^{\infty} a_n e^{in\theta} = e^{-ika \cos \theta}. \quad (5)$$

The problem is, therefore, simply that of determining the coefficients in the Fourier series representation of a function with period 2π . The answer is

$$a_n = (-i)^n J_n(ka). \quad (6)$$

Note that the method obviously works for any specified behavior of U^s on $r=a$.

Turn now to a consideration of the application of Fourier integral representations in diffraction theory. These representations arise most simply when the wave equation (1) is discussed in terms of two-dimensional rectangular Cartesian coordinates x, y . A separable solution is then

$$e^{iklz} e^{ikmy} \quad (7)$$

for any (possibly complex) numbers l and m such that

$$l^2 + m^2 = 1. \quad (8)$$

A rather general solution is therefore

$$\int p(l) e^{iklz} e^{ik\sqrt{1-l^2}y} dl, \quad (9)$$

where $p(l)$ is an arbitrary function of l (independent of x and y) and the integration is over some path in the complex l plane.

The solution (9) is evidently suitable for discussing the problem in which U^s is specified on the plane $y=0$. For if the path of integration is taken to be along the complete real l axis, then at $y=0$ (9) is

$$\int_{-\infty}^{\infty} p(l) e^{iklx} dl, \quad (10)$$

which can be matched to any function of x , $f(x)$ say, capable of being represented as a Fourier integral. The general mathematical statement is that $p(l)$ is $k/(2\pi)$ times the Fourier transform of $f(x)$; that is,

$$p(l) = \frac{k}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-iklx} dx. \quad (11)$$

With the path of integration in (9) along the real axis, and the natural convention that $\sqrt{1-l^2}$ is positive for $|l| < 1$, it is evident from the outgoing requirement that (9) is appropriate to the region $y < 0$ (provided this is source-free) and $\sqrt{1-l^2}$ must have a negative imaginary part when $|l| > 1$. The corresponding solution appropriate to the region $y > 0$ is that synthesized in a similar manner, with the separable solution

$$e^{iklz} e^{-ikmy} \quad (12)$$

replacing (7). Since, in the integration, l ranges from $-\infty$ to $+\infty$, there is evidently no gain in generality in considering the separable solutions obtained from (7) and (12) by changing the sign of l .

Two features of this simple discussion should be noted. First, the transform involved is that associated with the coordinate which varies along the specified boundary (in this case, x); second, the satisfaction of the outgoing radiation requirement is achieved by the correct choice of the part of the separable solution associated with the coordinate which is constant on the specified boundary (in this case, y).

A final and important point remains to be made. The analysis in terms of Fourier integrals is applicable even when the specified behavior on $y=0$ is periodic in x . Moreover, the integral representation preserves a flexibility, lost in the Fourier series representation, which permits the subsequent reduction to yield a form of solution suited to the particular conditions of the problem. For example, suppose that $f(x)$ in (11) is periodic in x , with the wavelength $2\pi/k'$. Then, certainly, one expression for $p(l)$, in terms of delta functions, is

$$p(l) = \sum_{n=-\infty}^{\infty} a_n \delta(l - nk'/k), \quad (13)$$

where a_n are the coefficients in the Fourier series expansion

$$f(x) = \sum_{n=-\infty}^{\infty} a_n e^{ik'n x}. \quad (14)$$

In this case the representation (9) of the field in $y < 0$ appears explicitly as a discrete spectrum of plane waves, each component of which is periodic in x , with a fundamental wavelength of $2\pi/k'$. On the other hand, $p(l)$ can evidently be represented as the sum of all functions

$$p(x_1, x_2; l) = \frac{k}{2\pi} \int_{x_1}^{x_2} f(x) e^{-iklx} dx \quad (x_2 > x_1) \quad (15)$$

in which the nonoverlapping ranges of integration $[x_1, x_2]$ together span the complete range $-\infty$ to ∞ . Such a representation, with suitable choice of the ranges $[x_1, x_2]$, may well yield a more tractable form of the field. This is obvious, for example, when the wavelength of $f(x)$ is much greater than both that of the radiation and the distance of the point of observation from the plane $y=0$ ($k' \ll k$, $k'y \ll 1$).

The physical interpretation of the analysis just given is clear, since the substitution of $p(x_1, x_2; l)$ for $p(l)$ in (9) obviously results in a field which on $y=0$ is $f(x)$ for $x_1 < x < x_2$, and is zero for either $x < x_1$ or $x > x_2$. The relation between the sum of such fields and the sum of the discrete plane waves associated with the representation (14) is closely analogous to the relation between the ray and mode formulations of the problem of a localized source in a waveguide.

The contention of the present paper is simply that the flexibility preserved by starting from an infinite integral transform, and its inverse, is sufficient to deal adequately with certain diffraction problems, irrespective of the dimensions of the body relative to the wavelength. In the circular cylinder problem, for example, a Fourier integral representation can immediately lead to a form of solution useful at short wavelengths in a way which seems to the writer both straightforward and illuminating. The analysis is, in fact, closely parallel to that of the simple discussion in Cartesian coordinates given in this section; physically, the argument is more subtle, because a consideration of "nonphysical" space is involved; but mathematically, the distinctions are only in matters of detail.

III. THE CIRCULAR CYLINDER

The problem, discussed in Section II, of the plane wave

$$U_i = e^{-ikr \cos \theta} \quad (16)$$

incident on the circular cylinder $r=a$, is here reconsidered in the light of the remarks subsequently made in Section II.

The separable solution of (1), analogous to (7), is

$$H_\nu^{(2)}(kr) e^{i\nu\theta}, \quad (17)$$

the distinction between (17) and (3) being that in (17) ν is not restricted to integer values. The general solution, analogous to (9), is

$$- \int_{-\infty}^{\infty} \frac{p(\nu)}{H_\nu^{(2)}(ka)} H_\nu^{(2)}(kr) e^{i\nu\theta} d\nu, \quad (18)$$

where the explicit factor $-1/H_\nu^{(2)}(ka)$ is included for convenience; and, since a Fourier integral is again in evidence, adequate generality is retained by taking the path of integration along the entire real ν axis. Moreover, the choice of the Hankel function of the second kind insures that (18) has the outgoing character at infinity requisite to any representation of the scattered field.

The solution for the scattered field is therefore given by (18), with $p(\nu)$ obtained by matching (18) to minus (16) on $r=a$. Thus

$$\int_{-\infty}^{\infty} p(\nu) e^{i\nu\theta} d\nu = e^{-ika \cos \theta}. \quad (19)$$

The right-hand side of (19) is a periodic function of θ , and the remarks made towards the end of Section II are relevant. It is convenient to write, formally,

$$p(\nu) = p(ka, \nu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ika \cos \theta} e^{-i\nu\theta} d\theta, \quad (20)$$

bearing in mind the different ways in which the right-hand side can be expressed.

Since it is not proposed to go into details here, attention will be confined to displaying one form of solution which, in general terms, is appropriate to the case $ka \gg 1$, and which thus contrasts with the classical solutions (5) and (6). To this end, appeal is made to Sommerfeld's integral representation of the Hankel functions, and the path of integration in (20) is distorted so that (20) is formally

$$p(ka, \nu) = \frac{1}{2} e^{-1/2 i \pi \nu} [H_\nu^{(1)}(ka) + H_\nu^{(2)}(ka)] \cdot \sum_{n=-\infty}^{\infty} e^{2i\pi n \nu}. \quad (21)$$

Then the total field, after being written

$$U = \int_{-\infty}^{\infty} \left\{ p(kr, \nu) - \frac{p(ka, \nu)}{H_\nu^{(2)}(ka)} H_\nu^{(2)}(kr) \right\} e^{i\nu\theta} d\nu, \quad (22)$$

appears, on the substitution of (21), as

$$U = \frac{1}{2} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \left[H_\nu^{(1)}(kr) - \frac{H_\nu^{(1)}(ka)}{H_\nu^{(2)}(ka)} H_\nu^{(2)}(kr) \right] \cdot e^{i(\theta + 2n\pi - 1/2\pi)\nu} d\nu. \quad (23)$$

The right-hand side of (23) is convergent, and the steps leading to it can easily be made rigorous. With $0 < \theta < \pi$,

for points in the shadow region the integrals are evaluated for all n by path closure around the zeros of $H_{\nu}^{(2)}(ka)$, and for points in the illuminated region they are evaluated by the same method, except that steepest descents must be used for the term $n=0$.

The key to arrival at (23) is, of course, the particular representation (21) of (20). This representation can be conceived as arising, in the first instance, from a division of the range of integration $[-\infty, \infty]$ for θ in (20) into the subranges $[-\pi/2 - 2n\pi, 3\pi/2 - 2n\pi]$, $n=0, \pm 1, \pm 2, \dots$. With the addition of arms joining the ends of the subranges to infinity in the appropriate regions in the lower half of the complex θ plane, the combined contributions of which cancel out, each subrange contributes to (21) the term with the corresponding value of n . The mathematical analogy with the analysis of the plane boundary problem of Section II is therefore evident. Stated physically, the terms in (23) with $n \neq 0$ are essentially concerned with the satisfaction of the boundary conditions in "non-physical" θ space, and it is, therefore, no surprise to find that they can be neglected when ka gets large enough.

IV. THE SPHERE

In this section is discussed the axially symmetrical problem of the plane wave

$$U^i = e^{-ikr \cos \theta} \quad (24)$$

incident on the sphere $r=a$, where r, θ , and ϕ are spherical polar coordinates ($0 \leq \theta < \pi$, $0 \leq \phi < 2\pi$).

Since the analysis is independent of ϕ , the expression of (1) in spherical polar coordinates leads to the separable solution

$$\frac{1}{\sqrt{r}} H_{\nu}^{(2)}(kr) E_{\nu-1/2}(\theta), \quad (25)$$

where the θ function is the solution of Legendre's equation (or order zero and degree $\nu-1/2$) defined in terms of a hypergeometric function by

$$E_{\nu-1/2}(\theta) = \sqrt{\frac{\pi}{2}} \frac{e^{i\pi/4}}{e^{i\pi/4}} \frac{(\nu - \frac{1}{2})!}{\nu!} \frac{e^{i\nu\theta}}{\sqrt{(\sin \theta)}} F\left(\frac{1}{2}, \frac{1}{2}; \nu + 1; -\frac{ie^{i\theta}}{2 \sin \theta}\right). \quad (26)$$

In (25) an attempt has been made to choose a solution analogous to (17). The function $E_{\nu-1/2}$ has, for example, the properties

$$E_{\nu-1/2}(\theta + m\pi) = i^m e^{i\pi m \nu} E_{\nu-1/2}(\theta), \quad m = 0, \pm 1, \pm 2, \dots \quad (27)$$

$$E_{\nu-1/2}(\theta) \sim \sqrt{\left(\frac{\pi}{2}\right)} \frac{e^{i\pi/4}}{\sqrt{(\nu \sin \theta)}}, \quad \text{as } |\nu| \rightarrow \infty \quad (\sin \theta \neq 0). \quad (28)$$

The corresponding general solution of (1) is therefore

$$-\sqrt{\left(\frac{a}{r}\right)} \int_{-\infty}^{\infty} \frac{p(\nu)}{H_{\nu}^{(2)}(ka)} H_{\nu}^{(2)}(kr) E_{\nu-1/2}(\theta) d\nu, \quad (29)$$

and to represent the scattered field, it only remains to choose $p(\nu)$ such that

$$\int_{-\infty}^{\infty} p(\nu) E_{\nu-1/2}(\theta) d\nu = e^{-ika \cos \theta}. \quad (30)$$

To solve (30), appeal is made to the infinite Legendre transform of a function $F(\theta)$, namely

$$f(\nu) = -\frac{i}{\pi} \int_{-\infty}^{\infty} F(\theta) E_{\nu-1/2}(\theta) \sin \theta d\theta. \quad (31)$$

and its inverse

$$F(\theta) = \frac{1}{\pi} \int_{-\infty}^{\infty} \nu f(\nu) \cot(\pi \nu) E_{\nu-1/2}(\theta) d\nu, \quad (32)$$

details of which are given by Clemmow.¹ Thus

$$p(\nu) \equiv p(ka, \nu) = -\frac{i}{\pi^2} \nu \cot(\pi \nu) \int_{-\infty}^{\infty} e^{-ika \cos \theta} E_{\nu-1/2}(\theta) \sin \theta d\theta. \quad (33)$$

The analysis then proceeds in a manner closely similar to that of Section III, and the total field appears in the form

$$U = -\frac{e^{i\pi/4}}{\sqrt{(2\pi)}} \frac{1}{\sqrt{(kr)}} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \nu \left[H_{\nu}^{(1)}(kr) - \frac{H_{\nu}^{(1)}(ka)}{H_{\nu}^{(2)}(ka)} H_{\nu}^{(2)}(kr) \right] \times e^{i(2n-1/2)\pi \nu} \cot(\pi \nu) E_{\nu-1/2}(\theta) d\nu. \quad (34)$$

V. THE PARABOLIC CYLINDER

In this section is discussed the two-dimensional problem of the plane wave

$$U^i = e^{ikr \cos(\theta-\alpha)} \quad (35)$$

incident on the parabolic cylinder $\eta = \sqrt{(2a)}$, where

$$\xi = \sqrt{(2r)} \cos\left(\frac{1}{2}\theta\right) \quad (-\infty < \xi < \infty) \quad (36)$$

and

$$\eta = \sqrt{(2r)} \sin\left(\frac{1}{2}\theta\right) \quad (0 < \eta < \infty) \quad (37)$$

are parabolic cylindrical coordinates (r and θ being the polar coordinates of Section III).

The appropriate separable solutions of (1) are

$$D_{-i\nu-1/2}[\pm \sqrt{(2k)} e^{i\pi/4} \xi] D_{i\nu-1/2}[\sqrt{(2k)} e^{i\pi/4} \eta], \quad (38)$$

in the notation standard for the parabolic cylinder

functions. It should be remarked that the choice of the η function is determined uniquely by the outgoing condition; but that two ξ functions (distinguished by the sign of the argument) must be considered, since one is *not* obtained from the other by reversing the sign of ν [see the remark following (12)].

The general solution which is to be particularized to give the scattered field is therefore

$$\begin{aligned} & \int_{-\infty}^{\infty} \{ p_1(\nu) D_{-i\nu-1/2} [\sqrt{(2k)} e^{i\pi/4} \xi] \\ & \quad + p_2(\nu) D_{-i\nu-1/2} [-\sqrt{(2k)} e^{i\pi/4} \xi] \} \\ & \quad \times D_{i\nu-1/2} [\sqrt{(2k)} e^{i\pi/4} \eta] d\nu. \end{aligned} \quad (39)$$

Now $p_1(\nu)$, $p_2(\nu)$ must be chosen so that the boundary conditions are satisfied; that is, on $\eta = \sqrt{(2a)}$, (39) must be minus (35). This is easily done by appealing to the infinite parabolic cylinder function transforms of a function $F(\xi)$, namely

$$\begin{aligned} f_1(\nu) &= \int_{-\infty}^{\infty} F(\xi) D_{i\nu-1/2} (\xi e^{-i\pi/4}) d\xi, \\ f_2(\nu) &= \int_{-\infty}^{\infty} F(\xi) D_{i\nu-1/2} (-\xi e^{-i\pi/4}) d\xi, \end{aligned} \quad (40)$$

and the inverse

$$\begin{aligned} F(\xi) &= \frac{1}{4\pi} \int_{-\infty}^{\infty} \frac{e^{\pi\nu/2}}{\cosh(\pi\nu)} \{ f_1(\nu) D_{-i\nu-1/2} (\xi e^{i\pi/4}) \\ & \quad + f_2(\nu) D_{-i\nu-1/2} (-\xi e^{-i\pi/4}) \} d\nu, \end{aligned} \quad (41)$$

relations established by Cherry.³

The result for the total field then turns out to be

$$\begin{aligned} U &= \frac{1}{2\sqrt{(\pi \sin \alpha)}} \int_{-\infty}^{\infty} \frac{e^{i\nu \log(\tan \alpha/2)}}{\cosh(\pi\nu)} \{ D_{i\nu-1/2} [-\sqrt{(2k)} e^{i\pi/4} \eta] \\ & \quad - \frac{D_{i\nu-1/2} [-2\sqrt{(ka)} e^{i\pi/4}]}{D_{i\nu-1/2} [2\sqrt{(ka)} e^{i\pi/4}]} D_{i\nu-1/2} [\sqrt{(2k)} e^{i\pi/4} \eta] \} \\ & \quad \cdot D_{-i\nu-1/2} [-\sqrt{(2k)} e^{i\pi/4} \xi] d\nu. \end{aligned} \quad (42)$$

Diffraction of Scalar Waves by a Circular Aperture*

J. BAZER AND A. BROWN†

Summary—The problems of the diffraction of scalar waves by a circular aperture in a perfectly soft, and in a perfectly rigid, infinite, planar screen are treated. New integral representations of the solution are presented which have the virtue of automatically satisfying the time-reduced wave equation, the radiation condition and the boundary conditions. The unknown functions appearing in these representations are shown to satisfy Fredholm-type integral equations of the second kind which yield, after iteration, accurate, approximate solutions when ka (the product of the wave number and the aperture radius) is sufficiently small. These approximate solutions are in turn employed to calculate the aperture fields, the far fields, the transmission coefficients, and the edge behaviors. The results are in complete agreement with known results and for some of these quantities they are more accurate in that higher powers of ka are included. Finally, in the case of the rigid screen problem, for all values of ka , we determine the form of the edge behavior and give a simple proof of the unique existence of the solution.

I. INTRODUCTION

WE are concerned with the problems of diffraction by a circular aperture in a perfectly soft and in a perfectly rigid, infinite, planar screen.¹ Our attention is focused mainly on the low frequency end of the spectrum, where the quantity $\alpha = ka$ (k is the wave number, a is the aperture radius) is small. We treat in detail only the case of normally-incident, plane-wave excitation although the method is applicable to arbitrary axially-symmetric excitation and may, in fact, be extended to cover arbitrary angularly-dependent excitation.² New integral representations $\mathcal{O}_1(f_1)$ and $\mathcal{O}_2(f_2)$ of the solutions of \mathcal{O}_1 and \mathcal{O}_2 are presented [f_1 and f_2 are unknown functions—see (11), (12) and (54)]. These representations have the virtue of being both simple and of leading directly to Fredholm integral equations of the second kind for f_1 and f_2 . We use the fact that the kernels of the Fredholm equations are small when α is small to calculate the first several terms of the Neumann series of f_1 and f_2 . These expressions are, in turn, employed to calculate approximate expressions for the aperture fields, the far fields, the transmission coefficients and the edge behaviors. The results thus obtained are in complete agreement with known results obtained by other means—such as have been summarized, for example, by C. J. Bouwkamp in his excellent review article [1]. Moreover, by including higher powers of α , we extend the accuracy of some of the results quoted by Bouwkamp. In addition, in \mathcal{O}_2 we give, for all values of $\alpha > 0$,

the form of the edge behavior and a simple proof of the unique³ existence of the solution.

The integral representations mentioned above form the core of our paper. They are the natural generalizations of similar well-known solutions of the potential-theoretic analogs of \mathcal{O}_1 and \mathcal{O}_2 such as may be found in the report [2] by S. N. Karp (our original source) and in the book [3]⁴ by Green and Zerna. Several authors, among whom are W. Magnus [4], N. Chako [5], D. S. Jones [6], N. I. Akhiezer and A. N. Akhiezer [7], B. Noble [8]⁵ and R. C. MacCamy and A. E. Heins [9],⁴ have succeeded in reducing aperture and disc problems to solving Fredholm integral equations of the second kind with kernels that are small when α is small.⁶ Our integral equations are most closely related to one first obtained by D. S. Jones [6].⁷ Although the general form of our inhomogeneous term is simpler, the kernels have fundamentally the same form. Mainly because of the simple form of the initial integral representations, our derivations of the Fredholm equations are shorter and more elementary than those of the authors mentioned above. In each of our representations, the field is related to the solution of the associated Fredholm equation by means of a single integral, and the derivation of the Fredholm equation involves only the inversion of a single convolution of the solution. In the work of Jones and of MacCamy and Heins,⁸ for example, double integrals replace the single integrals and successive convolutions replace the single convolutions.⁹ In addition, we can verify almost at a glance that the total fields satisfy the boundary conditions as well as the wave equations and the radiation condition; it is not necessary to revert to more primitive representations as in some of the above approaches. Furthermore, the general form of the fields near the edge can be ascertained by a simple calculation.

It is true that our approach depends essentially on guessing the proper form of the solution, whereas the several approaches mentioned above employ relatively familiar points of departure. It is our belief that the

* The research reported in this document has been sponsored by the AF Cambridge Research Center, Air Res. and Dev. Command, under Contract No. AF 19(604)5238.

† New York University, Inst. Math. Sci., New York, N. Y.

¹ The soft-screen problem, or first boundary-value problem, will hereafter be referred to as \mathcal{O}_1 and the rigid-screen problem, or second boundary-value problem, as \mathcal{O}_2 .

² This generalization will be treated in a forthcoming paper by J. Bazer and H. Hochstadt.

³ Establishing the uniqueness for both \mathcal{O}_1 and \mathcal{O}_2 is a relatively trivial matter [see discussion following (48) below].

⁴ We are indebted to B. Noble for bringing references [3] and [9] to our attention.

⁵ Using the so-called dual integral-equation approach, B. Noble finds it possible to treat both slit and aperture problems simultaneously.

⁶ We refer the reader to C. J. Bouwkamp's review [1] for a discussion of other methods.

⁷ In reference [6], Jones treats the rigid disk problem which, by virtue of Babinet's principle, is equivalent to our \mathcal{O}_1 .

⁸ The approach of these authors also leads to a Jones-type integral equation.

⁹ It is for this reason that the general form of the inhomogeneous terms of our Fredholm equations is simpler than that of Jones.

transparent nature of the representations and the elementary character of the subsequent analysis more than make up for this seeming defect. Moreover, preliminary calculations reveal that the scope of the method is by no means limited to scalar or to three-dimensional aperture problems. The same is undoubtedly true of the other approaches; however, the simplicity of the present method may well bring within reach orders of approximation that, for all practical purposes, are inaccessible to more cumbersome methods.

II. FORMULATION OF THE PROBLEM \mathcal{P}_1 AND \mathcal{P}_2

Let \mathcal{A} be a circular aperture of radius a in a rigid screen \mathcal{S} and let $\mathcal{A} + \mathcal{S}$ coincide with the (x, y) -plane, the origin being located at the center of \mathcal{A} (see Fig. 1).

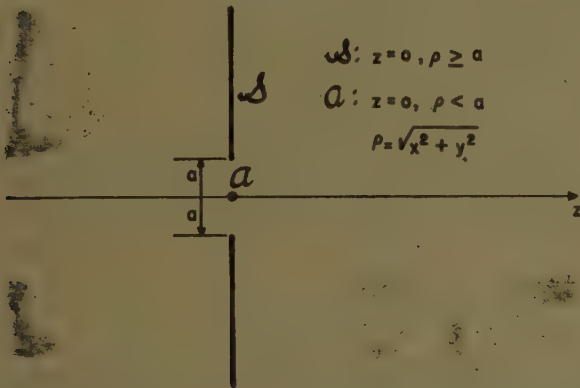


Fig. 1—A cross-sectional view of the diffraction screen. \mathcal{S} is the screen, \mathcal{A} is the aperture of radius a , centered at the origin of coordinates. $\mathcal{S} + \mathcal{A}$ is the (x, y) -plane.

Then the solution u_1 of the first boundary-value problem \mathcal{P}_1 may be expressed as follows [cf. reference [1], p. 38]:¹⁰

$$u_1(\rho, z) = \begin{cases} u^0(\rho, z) - u^0(\rho, -z) + \phi_1(\rho, -z), & z \leq 0, \\ \phi_1(\rho, z), & z \geq 0. \end{cases} \quad (1)$$

Here, ρ is defined by

$$\rho = \sqrt{x^2 + y^2}; \quad (2)$$

$u^0(\rho, z)$ denotes the axially-symmetric exciting wave incident from the left (in the region $z < 0$); $\phi_1(\rho, z)$ is defined in the half-space $z \geq 0$, and has the following properties:

- 1) ϕ_1 is a solution of the wave equation $\Delta\phi_1 + k^2\phi_1 = 0$, when $z > 0$;
- 2) $\phi_1 = 0$ on \mathcal{S} ;
- 3) ϕ_1 satisfies the Sommerfeld radiation condition at infinity;

$$4) \quad \frac{\partial\phi_1}{\partial z} = \frac{\partial u^0}{\partial z} \text{ in } \mathcal{A}; \quad (3)$$

- 5) ϕ_1 is everywhere finite;
- 6) $\nabla\phi_1$ is quadratically integrable over any domain in 3-space, including domains having some or all of the points of the circular edge as boundary points.

The corresponding solution u_2 of the second boundary-value problem \mathcal{P}_2 may be expressed as follows:

$$u_2(\rho, z) = \begin{cases} u^0(\rho, z) + u^0(\rho, -z) - \phi_2(\rho, -z), & z \leq 0, \\ \phi_2(\rho, z), & z \geq 0. \end{cases} \quad (4)$$

Here, $\phi_2(\rho, z)$ is defined for $z \geq 0$ and possesses all the properties listed above except 2) and 4), which must be replaced by

$$2)' \quad \frac{\partial\phi_2}{\partial z} = 0 \text{ on } \mathcal{S}; \quad (5)$$

$$4)' \quad \phi_2 = u^0 \text{ in } \mathcal{A}.$$

Let us write

$$u_1^0(\rho) = \frac{\partial u^0}{\partial z} \text{ in } \mathcal{A}, \quad (6)$$

and

$$u_2^0(\rho) = u^0 \text{ in } \mathcal{A}. \quad (7)$$

It is henceforth assumed that $u_j^0(\rho)$, $j = 1, 2$, is an even, analytic function of ρ whose analytic continuation is regular at least in a (closed) circular region of radius $a + \Delta$, $\Delta > 0$, centered at $\rho = 0$. It is also assumed that the Laplace transform of $u_j^0(\rho)$, $j = 1, 2$, ρ real, exists.¹¹ In the special case of normally incident, plane-wave excitation¹²

$$u^0 = e^{ikz}, \quad (8)$$

so that

$$u_1^0(\rho) = ik, \quad (9)$$

and

$$u_2^0(\rho) = 1. \quad (10)$$

These functions clearly have the required properties.

III. THE BASIC INTEGRAL REPRESENTATION FOR ϕ_1

We require that ϕ_1 have an integral representation of the form

$$\phi_1 = \int_{-1}^1 \frac{e^{ikR^+}(\rho, z; t)}{R^+(\rho, z; t)} f_1(t) dt. \quad (11)$$

Here, $f_1(t)$ is required to be an odd function of t , and

$$R^\pm(\rho, z; t) = \sqrt{\rho^2 + (z \pm iat)^2}, \quad (12)$$

where the upper and lower signs on the left correspond, respectively, to the upper and lower signs in the right

¹⁰ In view of the assumed axial symmetry of the incident excitation, all field quantities may be assumed to depend on $\rho = \sqrt{x^2 + y^2}$ and z alone.

¹¹ For sufficient conditions, we refer the reader to reference [10].

¹² Time dependence of the form $\exp(-i\omega t)$ is assumed throughout.

member. Evidently, (12) is equivalent to

$$\phi_1 = \int_0^1 \left\{ \frac{e^{ik} R^+(\rho, z; t)}{R^+(\rho, z; t)} - \frac{e^{ik} R^-(\rho, z; t)}{R^-(\rho, z; t)} \right\} f_1(t) dt. \quad (13)$$

Since the regularity of $u_1^0(\rho)$ in the (closed) region defined by $0 \leq |\rho| \leq a + \Delta$ (see the concluding paragraphs of Sec. III) ultimately implies that $f_1(t)$ is regular in the circular region \mathcal{C}_Δ defined by $0 \leq |t| \leq 1 + \Delta/a$, we may, without loss of generality, require $f_1(t)$ to be regular in \mathcal{C}_Δ at the outset. The branches of the square roots in (13) are fixed by requiring that

$$\lim_{z \downarrow 0} R^\pm(\rho, z; t) = + \sqrt{\rho^2 - a^2 t^2} \quad \text{when } \rho^2 > a^2 t^2. \quad (14)$$

This specification immediately implies that

$$\lim_{z \downarrow 0} R^\pm(\rho, z; t) = \pm i \sqrt{a^2 t^2 - \rho^2} \quad \text{when } \rho^2 < a^2 t^2. \quad (15)$$

At points in the aperture, the only points in physical space where the integrands of (11) and (12) may be singular, ϕ_1 and its first and higher order derivatives may be defined as the limits approached by ϕ_1 in $z > 0$ as z approaches zero. The actual limits can be derived with the aid of the regularity properties of $f_1(t)$ in \mathcal{C}_Δ and function-theoretic arguments such as are found in reference [11]. In Appendix I of our report [12], we give an illustration of the type of argument required. Here it is enough to say that the present limiting process and the several limiting processes employed in the sequel are justifiable and that the limits agree with those arrived at on the basis of purely formal considerations. Finally, from (11) we see that ϕ_1 may be described as the field resulting from a superposition of sources along the line segment $(-ia, ia)$, situated on the imaginary z axis.

Although the explicit form of $f(t)$ is as yet unknown, we can, nevertheless, verify that ϕ_1 satisfies all the conditions listed in (2) except 4), the aperture condition. That ϕ_1 satisfies property 1) follows directly from the fact that the functions $\exp [ik R^\pm(\rho, z; t)/R^\pm(\rho, z; t)]$ satisfy the wave equation in the region $z > 0$. To verify that ϕ_1 vanishes on the screen [property 2)], we need only let z approach zero in (13) and make use of the specified behavior of the square roots [see (14) and (15)]. To verify property 3), we first introduce the polar coordinates.

$$\begin{aligned} z &= R \cos \theta, \\ \rho &= R \sin \theta, \quad 0 \leq \theta \leq \pi/2, \\ R^2 &= \rho^2 + z^2 \end{aligned} \quad (16)$$

and the note that the behavior of $R^\pm(\rho, z; t)$ for large R is

$$R^\pm(\rho, z; t) = R \pm iat \cos \theta + O\left(\frac{1}{R}\right). \quad (17)$$

From this relation, it follows immediately that the Sommerfield radiation condition is satisfied. Since $\exp [ik R^\pm(\rho, z; t)/R^\pm(\rho, z; t)]$ and its derivatives with respect to ρ and z are continuous functions of t in the

half-space $z > 0$, we conclude that ϕ_1 and $\nabla \phi_1$ are finite in this region [see properties 5) and 6)]. To determine the behavior of ϕ_1 and $\nabla \phi_1$ near the edge, it is clearly sufficient in view of the axial symmetry, to fix on a given cross section. Let T_δ be a torus whose axis of symmetry is the z axis and whose equation in the cross section determined by the z axis and the point (ρ, z) is (see Fig. 2)

$$T_\delta; \begin{cases} z = \delta \sin \gamma, \\ \rho = a + \delta \cos \gamma, \end{cases} \quad 0 \leq \gamma \leq \pi. \quad (18)$$

Here, δ is the radius of the generating circle of T_δ and γ is the angle measured clockwise from the screen to the ray passing through the edge and the point (ρ, z) .

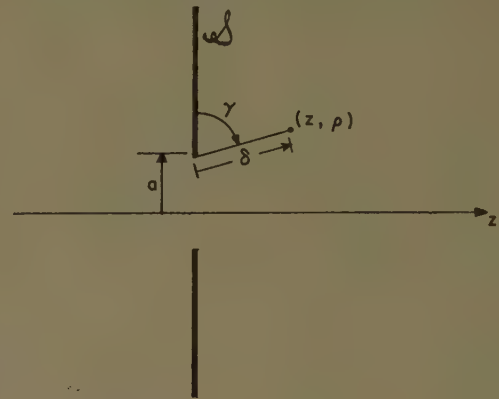


Fig. 2—A cross-sectional view of the diffracting screen. The angle γ is measured from the screen clockwise to the ray joining the edge of the screen to the point (z, ρ) ; δ is the distance from the edge to the point (z, ρ) .

Let us write

$$\hat{\phi}_1(\delta, \gamma) = \phi_1(a + \delta \cos \gamma, \delta \sin \gamma). \quad (19)$$

Then it can be shown (for a detailed derivation see Appendix I, reference [12]) that¹³

$$\hat{\phi}_1(\delta, \gamma) = \frac{-if_1(1)2\sqrt{2}}{a} (\delta/a)^{1/2} \sin(\gamma/2) + O(\delta/a), \quad (20)$$

and

$$\begin{aligned} (\delta/a)^{1/2} \frac{\partial \hat{\phi}_1(\delta, \gamma)}{\partial \delta} &= \frac{-i\sqrt{2}f_1(1)}{a^2} \sin(\gamma/2) + O[(\delta/a)^{1/2}], \end{aligned} \quad (21)$$

$$\begin{aligned} (\delta/a)^{1/2} \frac{1}{\delta} \frac{\partial \hat{\phi}_1(\delta, \gamma)}{\partial \gamma} &= \frac{-i\sqrt{2}f_1(1)}{a^2} \cos(\gamma/2) + O[(\delta/a)^{1/2}]. \end{aligned} \quad (22)$$

These equations and our earlier remarks show that ϕ_1 and $\nabla \phi_1$ satisfy the conditions 5) and 6).

¹³ For sufficiently small values of α , $f_1(1) = ia(a_1/2)$, where a_1 is a power series in α , the first several terms of which are given in (52).

Note that ϕ_1 and $\nabla\phi_1$ in the neighborhood of the edge have the same form as the corresponding field quantities in the neighborhood of a perfectly soft half-plane.

Letting z approach zero in (13) and using (14) and (15), we find that the aperture field reduces to

$$\phi_1(\rho, 0) = \frac{2}{ai} \int_{\rho/a}^1 \frac{\cosh[\alpha\sqrt{t^2 - (\rho/a)^2}]}{\sqrt{t^2 - (\rho/a)^2}} f_1(t) dt. \quad (23)$$

Moreover, setting

$$A_1(\theta) = -2 \int_0^1 \sinh(\alpha t \cos \theta) f_1(t) dt, \quad (24)$$

we readily verify that the far field of ϕ_1 and the transmission coefficient t_1 can be expressed as follows (cf. reference [1], p. 71):

$$\phi_1 \sim A_1(\theta) \frac{e^{ikR}}{R}, \quad (25)$$

$$t_1 = \frac{2}{a^2} \int_0^{\pi/2} |A_1(\theta)|^2 \sin \theta d\theta. \quad (26)$$

Approximate expressions for these quantities associated with normally-incident, plane-wave excitation are given in Sec. IV and V.

IV. DERIVATION OF THE INTEGRAL EQUATION FOR $f_1(t)$

The unknown function $f_1(t)$ may be determined by enforcing the aperture condition 4) of (3) which, by virtue of equation (10), may be expressed as follows:

$$\lim_{z \downarrow 0} \frac{\partial \phi_1(\rho, z)}{\partial z} = \frac{\partial u^0(\rho, 0)}{\partial z} \equiv u_1^0(\rho), \quad 0 \leq \rho < 1. \quad (27)$$

We begin by calculating the limiting value of $\partial\phi_1/\partial z$. Assuming that $z > 0$, we differentiate both sides of equation (13), and obtain the following relation:

$$\begin{aligned} \frac{\partial \phi_1}{\partial z} = & \frac{1}{\rho} \frac{\partial}{\partial \rho} \int_0^1 \frac{(z + ial) \exp[ikR^+(\rho, z; t)]}{R^+(\rho, z; t)} f_1(t) dt \\ & - \frac{1}{\rho} \frac{\partial}{\partial \rho} \int_0^1 \frac{(z - ial) \exp[ikR^-(\rho, z; t)]}{R^-(\rho, z; t)} f_1(t) dt. \end{aligned} \quad (28)$$

If we now let z approach zero and make use of the limiting behavior of $R^\pm(\rho, z; t)$ when $0 < \rho^2 \leq a^2 t^2$ and $a^2 t^2 \leq \rho^2 \leq a^2$, $0 \leq t \leq 1$ [see (14) and (15)], we find that for all ρ , $0 < \rho < a$,

$$\begin{aligned} \lim_{z \downarrow 0} \left[\frac{\partial \phi_1(\rho, z)}{\partial z} \right] = & \frac{i}{\rho} \frac{\partial}{\partial \rho} \int_0^{\rho/a} \left\{ \frac{\exp[i\alpha\sqrt{(\rho/a)^2 - t^2}]}{\sqrt{(\rho/a)^2 - t^2}} \right. \\ & \left. + \frac{\exp[i\alpha\sqrt{(\rho/a)^2 - t^2}]}{\sqrt{(\rho/a)^2 - t^2}} \right\} t f_1(t) dt \\ & + \frac{1}{\rho} \frac{\partial}{\partial \rho} \int_{\rho/a}^1 \left\{ \frac{\exp[-\alpha\sqrt{t^2 - (\rho/a)^2}]}{\sqrt{t^2 - (\rho/a)^2}} \right. \\ & \left. - \frac{\exp[\alpha\sqrt{t^2 - (\rho/a)^2}]}{\sqrt{t^2 - (\rho/a)^2}} \right\} t f_1(t) dt. \end{aligned} \quad (29)$$

Hence,

$$\begin{aligned} \lim_{z \downarrow 0} \left[\frac{\partial \phi_1}{\partial z} \right] = & \frac{2i}{\rho} \frac{\partial}{\partial \rho} \int_0^{\rho/a} \frac{\cos(\alpha\sqrt{(\rho/a)^2 - t^2})}{\sqrt{(\rho/a)^2 - t^2}} t f_1(t) dt \\ & - \frac{2}{\rho} \frac{\partial}{\partial \rho} \int_0^{\rho/a} \frac{\sin(\alpha\sqrt{(\rho/a)^2 - t^2})}{\sqrt{(\rho/a)^2 - t^2}} t f_1(t) dt \\ & - \frac{2}{\rho} \frac{\partial}{\partial \rho} \int_{\rho/a}^1 \frac{\sinh(\alpha\sqrt{t^2 - (\rho/a)^2})}{\sqrt{t^2 - (\rho/a)^2}} t f_1(t) dt. \end{aligned} \quad (30)$$

Since

$$\sin(\alpha\sqrt{(\rho/a)^2 - t^2})/\sqrt{(\rho/a)^2 - t^2}, \quad t^2 \leq (\rho/a)^2,$$

is the continuation of

$$\sinh(\alpha\sqrt{(\rho/a)^2 - t^2})/\sqrt{(\rho/a)^2 - t^2}, \quad (\rho/a)^2 < t^2,$$

we conclude that for all ρ , $0 < \rho < a$,

$$\begin{aligned} \lim_{z \downarrow 0} \left[\frac{\partial \phi_1(\rho, z)}{\partial z} \right] = & \frac{2i}{\rho} \frac{\partial}{\partial \rho} \int_0^{\rho/a} \frac{\cos(\alpha\sqrt{(\rho/a)^2 - t^2})}{\sqrt{(\rho/a)^2 - t^2}} t f_1(t) dt \\ & - \frac{2}{\rho} \frac{\partial}{\partial \rho} \int_0^1 \frac{\sinh(\alpha\sqrt{t^2 - (\rho/a)^2})}{\sqrt{t^2 - (\rho/a)^2}} t f_1(t) dt, \end{aligned} \quad (31)$$

where it must be understood that

$$\sqrt{t^2 - (\rho/a)^2} = -i\sqrt{(\rho/a)^2 - t^2} \text{ when } t^2 < (\rho/a)^2. \quad (32)$$

If the value at $\rho=0$ of the right member of (31) is defined to be the limit approached as ρ approaches zero, then it can be shown that (31) is valid for all ρ in the half-open interval $0 \leq \rho < a$.

Invoking the aperture condition (27), we find that

$$\begin{aligned} \frac{\partial}{\partial \rho} \int_0^{\rho/a} \frac{\cos(\alpha\sqrt{(\rho/a)^2 - t^2})}{\sqrt{(\rho/a)^2 - t^2}} t f_1(t) dt \\ = \frac{\rho u_1^0(\rho)}{2i} + \frac{1}{i} \frac{\partial}{\partial \rho} \int_0^1 \frac{\sinh[\alpha\sqrt{t^2 - (\rho/a)^2}]}{\sqrt{t^2 - (\rho/a)^2}} t f_1(t) dt, \end{aligned} \quad 0 \leq \rho < a. \quad (33)$$

Integrating both sides from 0 to ρ , we obtain the integral equation

$$\begin{aligned} \int_0^{\rho/a} \frac{\cos(\alpha\sqrt{(\rho/a)^2 - t^2})}{\sqrt{(\rho/a)^2 - t^2}} t f_1(t) dt = & \frac{1}{2i} \int_0^{\rho} \xi u_1^0(\xi) d\xi \\ & + \frac{1}{i} \int_0^1 \frac{\sinh(\alpha\sqrt{t^2 - (\rho/a)^2})}{\sqrt{t^2 - (\rho/a)^2}} t f_1(t) dt - C, \end{aligned} \quad (34)$$

where

$$C = \frac{1}{i} \int_0^1 \frac{\sinh \alpha s}{s} [s f_1(s)] ds. \quad (35)$$

When we introduce the new variables,

$$\xi = t^2 \quad (36)$$

and

$$\eta = (\rho/a)^2, \quad (37)$$

into both sides of (34), this equation assumes the following form:

$$\int_0^\eta \frac{\cos(\alpha\sqrt{\eta-\xi})}{\sqrt{\eta-\xi}} F(\xi) d\xi = G(\eta) + H(\eta), \quad (38)$$

where

$$F(\xi) = f(\sqrt{\xi})/2, \quad (39)$$

$$G(\eta) = \frac{1}{2i} \int_0^{a\sqrt{\eta}} \xi u_1^0(\xi) d\xi, \quad (40)$$

$$H(\eta) = \frac{1}{i} \int_0^1 \frac{\sinh(\alpha\sqrt{\xi-\eta})}{\sqrt{\xi-\eta}} F(\xi) d\xi - C. \quad (41)$$

Using the fact that the Laplace transformation of a convolution of two functions is the product of the Laplace transforms, we may solve equation (38) for $F(\xi)$. The final result is

$$F(\eta) = \frac{1}{\pi} \frac{d}{d\eta} \int_0^\eta \frac{\cosh \alpha\sqrt{\eta-\mu}}{\sqrt{\eta-\mu}} [G(\mu) + H(\mu)] d\mu \\ = g(\eta) + h(\eta), \quad (42)$$

where $g(\eta)$ and $h(\eta)$ are defined by

$$g(\eta) = \frac{1}{\pi} \frac{d}{d\eta} \int_0^\eta \frac{\cosh(\alpha\sqrt{\eta-\xi})}{\sqrt{\eta-\xi}} G(\xi) d\xi \\ = \frac{1}{2\pi i} \frac{d}{d\eta} \int_0^\eta \frac{\cosh(\alpha\sqrt{\eta-\xi})}{\sqrt{\eta-\xi}} \\ \times \left(\int_0^{a\sqrt{\xi}} \xi u_1^0(\xi) d\xi \right) d\xi, \quad (43)$$

$$h(\eta) = \frac{1}{\pi} \frac{d}{d\eta} \int_0^\eta \frac{\cosh(\alpha\sqrt{\eta-\mu})}{\sqrt{\eta-\mu}} H(\mu) d\mu \\ = \frac{1}{\pi i} \int_0^1 F(\xi) \\ \cdot \left[\frac{d}{d\eta} \int_0^\eta \frac{\cosh(\alpha\sqrt{\eta-\mu}) \sinh(\alpha\sqrt{\eta-\mu})}{\sqrt{\eta-\mu} \sqrt{\xi-\mu}} d\mu \right] d\xi \\ - \frac{C}{\pi} \cosh \alpha\sqrt{\eta}. \quad (44)$$

The expression for $g(\eta)$ may be simplified somewhat, first by integrating by parts and then by making the change of variables $\xi = \eta - \eta s^2$. The result is:

$$g(\eta) = \frac{a^2 \sqrt{\eta}}{2\pi i} \int_0^1 \cosh(\alpha\sqrt{\eta}s) u_1^0[a\sqrt{\eta}(1-s^2)] ds. \quad (45)$$

The expression for $h(\eta)$ may also be simplified¹⁴ by 1), making use of the addition formula for hyperbolic functions, 2) introducing obvious changes of variables, and 3) carrying out the differentiation with respect to η . The final result is

$$h(\eta) = \frac{1}{2\pi i} \int_0^1 F(\xi) \\ \cdot \left\{ \frac{\sinh[\alpha(\sqrt{\xi}-\sqrt{\eta})]}{\sqrt{\eta}(\sqrt{\xi}-\sqrt{\eta})} + \frac{\sinh[\alpha(\sqrt{\xi}+\sqrt{\eta})]}{\sqrt{\eta}(\sqrt{\xi}+\sqrt{\eta})} \right\} d\xi \\ - C \cosh(\alpha\sqrt{\eta})/\pi. \quad (46)$$

Combining (42)–(46), we find

$$F(\eta) = \frac{a^2 \sqrt{\eta}}{2\pi i} \int_0^1 \cosh(\alpha\sqrt{\eta}s) u_1^0[a\sqrt{\eta}(1-s^2)] ds \\ + \frac{1}{2\pi i} \int_0^1 F(\xi) \\ \cdot \left\{ \frac{\sinh \alpha(\sqrt{\xi}-\sqrt{\eta})}{\sqrt{\eta}(\sqrt{\xi}-\sqrt{\eta})} + \frac{\sinh \alpha(\sqrt{\xi}+\sqrt{\eta})}{\sqrt{\eta}(\sqrt{\xi}+\sqrt{\eta})} \right\} d\xi \\ - C \cosh(\alpha\sqrt{\eta})/\pi. \quad (47)$$

Replacing η by t^2 and ξ by s^2 wherever ξ and η appear in (47) and taking into account 1) the definitions of C [(35)] and $F(\xi)$ [(39)] and 2) the fact that $f_1(t)$ is an odd function of t , we obtain the following integral equation for the function $tf_1(t)$:

$$tf_1(t) = \frac{a^2 t^2}{\pi i} \int_0^1 \cosh(\alpha ts) u_1^0[at\sqrt{(1-s^2)}] ds \\ + \frac{1}{\pi i} \int_{-1}^1 \left\{ \frac{\sinh[\alpha(t-s)]}{t-s} \right. \\ \left. - \frac{\cosh(\alpha t) \sinh(\alpha s)}{s} \right\} sf_1(s) ds. \quad (48)$$

In the special case of normal plane-wave incidence, where $u_1^0(\rho) = ik$ [see (10) and (27)], (48) becomes

$$tf_1(t) = \frac{at}{\pi} \sinh \alpha t + \frac{1}{\pi i} \int_{-1}^1 \left\{ \frac{\sinh[\alpha(t-s)]}{t-s} \right. \\ \left. - \frac{\cosh(\alpha t) \sinh(\alpha s)}{s} \right\} sf_1(s) ds. \quad (49)$$

It is easily verified that the kernel in the integral equation (48) is of order α^3 when α is small. Thus, assuming merely that $u_1^0(\rho)$, $\rho > 0$, is a continuous function of ρ , we readily conclude that equation (48) has a unique continuous solution (the Neumann series) for sufficiently small values of α .¹⁵ With this fact established, it readily

¹⁴ Cf. reference [6] pp. 10–11.

¹⁵ In problem \mathcal{P}_1 we are able to establish the unique existence of $f_2(t)$ for all $\alpha > 0$ because the kernel $\sinh[\alpha(t-s)]/t-s$ [see (64) and (66)] is symmetric and because the complex constant $\lambda = 1/\pi i$ can never be an eigenvalue of the corresponding homogeneous integral equation.

follows that the domain of regularity of $tf_1(t)$ contains the circular domain \mathcal{C}_Δ defined by the relation $0 \leq |t| \leq 1 + \Delta/a$. To prove this, it is enough to observe that 1) the integral term is an entire function of t since the kernel enjoys this property and 2) the inhomogeneous term is regular for all t in the domain \mathcal{C}_Δ because $u_1^0(\rho)$ is, by hypothesis, a regular function of ρ in the circular domain $0 \leq \rho \leq a + \Delta$. In effect, the regularity properties of $u_1^0(\rho)$ control the regularity properties of $tf_1(t)$. That $tf_1(t)$ is even follows from the form of the kernel and the fact that the inhomogeneous term is even—it is here that we make use of the evenness of $u_1^0(\rho)$ as a function of ρ . Finally, let us note that $f_1(t) = O(e^{\alpha|t|})$ when $|t|$ becomes large so that its Laplace transform exists.¹⁶

Assuming that ϕ_1 exists, it remains to prove that ϕ_1 is unique in the class of solutions satisfying the properties listed in (3). To this end we employ the standard argument which, briefly, runs as follows. If ϕ_1' is another solution, then $v = \phi_1 - \phi_1'$ enjoys all the properties of ϕ_1 except that $\partial v / \partial z$ vanishes in the aperture. We next consider the integral $\int_{\mathcal{D}} \nabla(v^* \cdot \nabla v) dV$; here, v^* denotes the complex conjugate of v , \mathcal{D} is the domain contained within the surface shown (in cross section) in Fig. 3 and dV is the volume element.

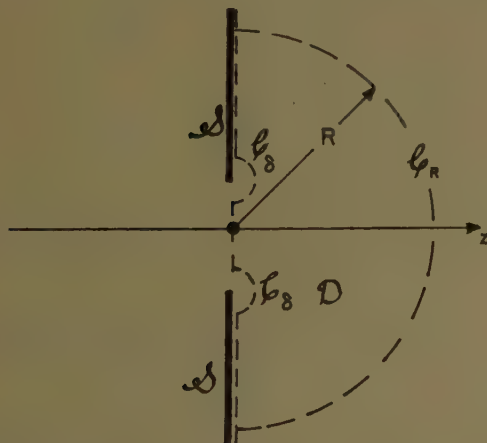


Fig. 3—A cross-sectional view of the surfaces (shown dashed) that bound the domain \mathcal{D} . \mathcal{C}_R and \mathcal{C}_a are meridional semicircles cut out of a hemisphere and torus whose equatorial planes coincide with the plane of the screen S . Both torus and hemisphere are centered at 0.

Employing 1) Green's Theorem, 2) the vanishing of v and $\partial v / \partial z$ on the aperture and the screen, respectively, and 3) the fact that v satisfies the wave equation and the radiation condition, we find, on letting ϵ and R^{-1} approach zero, that $\int_{\mathcal{D}} (|\nabla v|^2 + k^2 |v|^2) dV = 0$. From this, it follows easily that v vanishes everywhere; i.e., that $\phi_1 = \phi_1'$.

Turning now to the special case of normal, plane-wave incidence, we use the fact that the kernel is of the order α^3 when α is small, and calculate several terms of

the Neumann series for $tf_1(t)$. We find after three iterations that $f_1(t)$ ¹⁷ is

$$f_1(t) = \frac{a}{\pi} \left\{ \left[\alpha t + \frac{\alpha^3 t^3}{6} + \frac{\alpha^5 t^5}{120} + \frac{\alpha^7 t^7}{5040} + \frac{\alpha^9 t^9}{362880} \right. \right. \\ \left. - \frac{4}{\pi^2} \left(\left[\frac{\alpha^7}{81} + \frac{7\alpha^9}{2025} \right] t + \frac{\alpha^9}{810} t^3 \right) \right] \\ + \frac{2i\alpha}{\pi} \left[\left(\left[\frac{\alpha^3}{9} + \frac{4\alpha^5}{225} + \frac{\alpha^7}{735} + \frac{8\alpha^9}{127575} \right] t \right. \right. \\ \left. + \left[\frac{\alpha^5}{90} + \frac{\alpha^7}{525} + \frac{31\alpha^9}{198450} \right] t^3 \right. \\ \left. + \left[\frac{\alpha^7}{2520} + \frac{\alpha^9}{14175} \right] t^5 + \frac{\alpha^9}{136080} t^7 \right) \\ \left. - \frac{4}{\pi^2} \left(\frac{\alpha^9}{729} \right) t \right] \right\} + O(\alpha^{11}). \quad (50)$$

Inserting this expression for $f_1(t)$ into (23), (24) and (26), we find that the aperture field is

$$\phi_1 = \sum_{n=1}^{\infty} a_n [1 - (\rho^2/a^2)]^{n-1/2}, \quad (51)$$

where

$$a_1 = \frac{2\alpha}{\pi i} \left[1 + \frac{\alpha^2}{6} + \frac{2i\alpha^3}{9\pi} + \frac{\alpha^4}{120} + \frac{13i\alpha^5}{225\pi} \right. \\ \left. + \left(\frac{1}{5040} - \frac{4}{81\pi^2} \right) \alpha^6 + \frac{323i\alpha^7}{44100\pi} \right. \\ \left. + \left(\frac{1}{362880} - \frac{38}{2025\pi^2} \right) \alpha^8 \right. \\ \left. + \left(\frac{269i}{5715360\pi} - \frac{4i}{729\pi^3} \right) \alpha^9 \right] + O(\alpha^{11}), \\ a_2 = \frac{\alpha^3}{\pi i} \left[1 + \frac{3\alpha^2}{10} + \frac{2i\alpha^3}{5\pi} + \frac{\alpha^4}{56} + \frac{19i\alpha^5}{175\pi} \right. \\ \left. + \left(\frac{1}{2160} - \frac{4}{45\pi^2} \right) \alpha^6 + \frac{631i\alpha^7}{396900} \right] + O(\alpha^{11}), \\ a_3 = \frac{\alpha^5}{300\pi i} \left[1 + \frac{5\alpha^2}{14} + \frac{10i\alpha^3}{21\pi} + \frac{5\alpha^4}{216} + \frac{76i\alpha^5}{567\pi} \right] + O(\alpha^{11}), \\ a_4 = \frac{\alpha^7}{17640\pi i} \left[1 + \frac{7\alpha^2}{18} + \frac{58\alpha^3 i}{81\pi} \right] + O(\alpha^{11}), \\ a_5 = \frac{\alpha^9}{(1632960)\pi i} + O(\alpha^{11}). \quad (52)$$

¹⁷ The maximum value of α for which this expression for $f_1(t)$, regarded as a power series in α , converges is at present unknown.

¹⁶ This justifies our earlier use of Laplace transform methods.

The diffracted amplitude is:

$$\begin{aligned}
 A_1(\theta) = & \frac{-2a}{\pi} \left\{ \left[\left(\frac{\alpha^2}{3} + \frac{\alpha^4}{30} + \frac{\alpha^6}{840} + \frac{\alpha^8}{45360} + \frac{\alpha^{10}}{3991680} \right) \cos \theta \right. \right. \\
 & + \left(\frac{\alpha^4}{30} + \frac{\alpha^6}{252} + \frac{\alpha^8}{6480} + \frac{\alpha^{10}}{332640} \right) \cos^3 \theta \\
 & + \left(\frac{\alpha^6}{840} + \frac{\alpha^8}{6480} + \frac{\alpha^{10}}{158400} \right) \cos^5 \theta \\
 & + \left(\frac{\alpha^8}{45360} + \frac{\alpha^{10}}{332640} \right) \cos^7 \theta + \left(\frac{\alpha^{10}}{3991680} \right) \cos^9 \theta \left. \right] \\
 & - \frac{4}{\pi^2} \left[\left(\frac{\alpha^8}{243} + \frac{17\alpha^{10}}{12150} \right) \cos \theta + \frac{\alpha^{10}}{2430} \cos^3 \theta \right] \\
 & + \frac{2i\alpha}{\pi} \left[\left(\frac{\alpha^4}{27} + \frac{11\alpha^6}{1350} + \frac{131\alpha^8}{147000} \right) \cos \theta \right. \\
 & + \left(\frac{\alpha^6}{270} + \frac{3\alpha^8}{3500} \right) \cos^3 \theta + \left. \frac{\alpha^8}{7560} \cos^5 \theta \right] \left. \right\} \\
 & + O(\alpha^{11}). \quad (53)
 \end{aligned}$$

The transmission coefficient is

$$\begin{aligned}
 t_1 = & \frac{8\alpha^4}{27\pi^2} \left[1 + \frac{8\alpha^2}{25} + \frac{311\alpha^4}{6125} + \left(\frac{2612}{496125} - \frac{4}{81\pi^2} \right) \alpha^6 \right] \\
 & + O(\alpha^{12}). \quad (54)
 \end{aligned}$$

These expressions are in complete agreement with those given by C. J. Bouwkamp (see reference [1], pp. 64 and 72). Actually, our result for the aperture field is a little more accurate than that given by Bouwkamp; we have calculated a_1, a_2, \dots, a_5 , regarded as power series in α , up to and including α^{10} ; Bouwkamp gives a_1, \dots, a_4 up to and including α^8 .

This completes our discussion of the first boundary-value problem.

V. THE SECONDARY BOUNDARY-VALUE PROBLEM

We represent $\phi_2(\rho, z)$ as follows:

$$\begin{aligned}
 \phi_2(\rho, z) = & \int_{-1}^1 \frac{e^{ikR^+(\rho, z; t)}}{R^+(\rho, z; t)} f_2(t) dt \\
 = & \int_0^1 \left\{ \frac{e^{ikR^+(\rho, z; t)}}{R^+(\rho, z; t)} + \frac{e^{ikR^-(\rho, z; t)}}{R^-(\rho, z; t)} \right\} f_2(t) dt. \quad (55)
 \end{aligned}$$

The quantities $R^\pm(\rho, z; t)$ are defined in (13); $f_2(t)$ is required to be 1) an even function of t and 2) a regular function of t in a circular domain of the complex t plane containing the unit disc. At points of the aperture, the only points in physical space where the integrand can be singular, ϕ_2 and its derivatives may be defined as the limits approached by ϕ_2 in $z > 0$ as z approaches zero.¹⁸

¹⁸ In the expression for ϕ_1 [(31)], the pole singularity of the integrand associated with $\rho = z = 0$ is only apparent since $f_1(t)$ is odd. On the other hand, the integrand of ϕ_2 at $\rho = z = 0$ has a genuine pole. In this case, $\lim_{z \rightarrow 0} \phi_2(0, z)$ may, for example, be obtained from the well-known Plemelj formulas $z \downarrow 0$ (see reference [11]).

Repeating the argument of Sec. III, we find that conditions 1), 2) and 3) [see Sec. II] are satisfied. In Appendix I of reference [12], we show that ϕ_2 and $\nabla \phi_2$ have the following behavior in the neighborhood of the edge:¹⁹

$$\hat{\phi}_2 = 1 - \frac{2\sqrt{2}f_2(1)}{a} (\delta/a)^{1/2} \cos(\gamma/2) + O(\delta/a), \quad (56)$$

$$(\delta/a)^{1/2} \frac{\partial \hat{\phi}_2}{\partial \delta} = \frac{-\sqrt{2}f_2(1)}{a^2} \cos(\gamma/2) + O[(\delta/a)^{1/2}], \quad (57)$$

$$(\delta/a)^{1/2} \frac{1}{\delta} \frac{\partial \hat{\phi}_2}{\partial \gamma} = \frac{\sqrt{2}f_2(1)}{a^2} \sin(\gamma/2) + O[(\delta/a)^{1/2}]. \quad (58)$$

The variables δ and γ are those introduced earlier in (18) (see Fig. 2) and $\hat{\phi}_2 = \hat{\phi}(\delta, \gamma)$ is defined by

$$\hat{\phi}_2 = \phi_2(\delta \cos \gamma, a + \delta \sin \gamma). \quad (59)$$

These results imply that conditions 5) and 6) of Sec. II are satisfied.

Employing (14), (15) and (17), we obtain from (55) the following expressions for $\partial \phi_2 / \partial z$ (in the aperture), the far field and the transmission coefficient t_2 :

$$\begin{aligned}
 \frac{\partial \phi_2}{\partial z} = & \frac{2}{\rho} \frac{\partial}{\partial \rho} \int_{\rho/a}^1 \frac{\cosh[\alpha\sqrt{t^2 - (\rho/a)^2}]}{\sqrt{t^2 - (\rho/a)^2}} t f_2(t) dt, \\
 & 0 \leq \rho < a, \quad (60)
 \end{aligned}$$

$$\phi_2 \sim A_2(\theta) \frac{e^{ikR}}{R}, \quad R \rightarrow \infty, \quad (61)$$

where

$$A_2(\theta) = 2 \int_0^1 \cosh(\alpha t \cos \theta) f_2(t) dt; \quad (62)$$

$$t_2 = \frac{2}{a^2} \int_0^{\pi/2} |A_2(\theta)|^2 \sin \theta d\theta. \quad (63)$$

Enforcing the aperture condition, we find the following analog of (34):

$$\begin{aligned}
 & \int_0^{\rho/a} \frac{\cos[\alpha\sqrt{(\rho/a)^2 - t^2}]}{\sqrt{(\rho/a)^2 - t^2}} f_2(t) dt \\
 = & \frac{a}{2} u_2^0(\rho) + \frac{1}{i} \int_0^1 \frac{\sinh(\alpha\sqrt{t^2 - (\rho/a)^2})}{\sqrt{t^2 - (\rho/a)^2}} \\
 & \times f_2(t) dt. \quad (64)
 \end{aligned}$$

Arguing as in Section IV, we are led to the integral equation,

$$f_2(t) = g(t) + \frac{1}{\pi i} \int_{-1}^1 \frac{\sinh[\alpha(t-s)]}{t-s} f_2(s) ds, \quad (65)$$

¹⁹ $f_2(1) = (a/\pi)a_0$, where a_0 is defined in (69) below. Several terms of the power series for a_0 , regarded as a function of α , are given in (70).

where

$$g(t) = \frac{a}{\pi} u_2^0(0) \cosh \alpha t + \frac{a^2 t}{\pi} \int_0^1 \cosh \alpha t \frac{u_2^{0'}(at\sqrt{1-s^2}) ds}{\sqrt{1-s^2}} \quad (66)$$

$$\left(u_2^{0'}(\rho) = \frac{d}{d\rho} u_2^0(\rho) \right).$$

In the special case of normally incident, plane-wave excitation of unit amplitude, we find, since $u_2^{0'}$ vanishes, that

$$f_2(t) = \frac{a}{\pi} \cosh \alpha t + \frac{1}{\pi i} \int_{-1}^1 \sinh \frac{[\alpha(t-s)]}{(t-s)} f_2(s) ds. \quad (67)$$

Eq. (65) implies the unique existence of an even function of t , regular in the domain $|t| \leq 1 + \Delta/a$, if, as we have assumed, $u_2^0(\rho)$ is an even function of ρ , regular in the domain $0 \leq |\rho| \leq a + \Delta$. Arguments paralleling those employed in Sec. IV guarantee the existence of the Laplace transform of $f_2(\sqrt{\xi})/\sqrt{\xi}$ so that the manipulation leading to (65)–(67) is justified. Making use of the fact that the kernel in (65) or (67) is of order α when α is small, we may obtain f as a Neumann series. In the plane-wave case [(67)], we find that²⁰

$$f_2(t) = \frac{a}{\pi} \left\{ \left[1 + \frac{\alpha^2 t^2}{2} + \frac{\alpha^4 t^4}{24} + \frac{\alpha^6 t^6}{720} \right] - \frac{i}{\pi} \left[\left(2\alpha + \frac{4\alpha^3}{9} + \frac{4\alpha^5}{75} \right) + \left(\frac{\alpha^3}{3} + \frac{4\alpha^5}{45} \right) t^2 + \frac{\alpha^5}{60} t^4 \right] - \frac{1}{\pi^2} \left[\left(4\alpha^2 + \frac{4\alpha^4}{3} + \frac{508\alpha^6}{2025} \right) + \left(\frac{2\alpha^4}{3} + \frac{34\alpha^6}{135} \right) t^2 + \frac{\alpha^6}{30} t^4 \right] + \frac{i}{\pi^3} \left[\left(8\alpha^3 + \frac{32\alpha^5}{9} \right) + \frac{4\alpha^5}{3} t^2 \right] + \frac{1}{\pi^4} \left[\left(16\alpha^4 + \frac{80\alpha^6}{9} \right) + \frac{8\alpha^6}{3} t^2 \right] - \frac{i}{\pi^5} [32\alpha^5] - \frac{1}{\pi^6} [64\alpha^6] \right\} + O(\alpha^7). \quad (68)$$

Inserting this expression into (60), (62) and (63), we obtain the following expressions for $\partial\phi_2/\partial z$, $A_2(\theta)$ and t_2 :

$$\frac{\partial\phi_2}{\partial z} = -\frac{2}{a\pi} [a_0[1 - (\rho/a)^2]^{-1/2} + a_1[1 - (\rho/a)^2]^{1/2} + a_2[1 - (\rho/a)^2]^{3/2} + a_3[1 - (\rho/a)^2]^{5/2}] + O(\alpha^7), \quad (69)$$

where

$$a_0 = 1 - \frac{2i}{\pi} \alpha + \left(\frac{1}{2} - \frac{4}{\pi^2} \right) \alpha^2 - i \left(\frac{7}{9\pi} - \frac{8}{\pi^3} \right) \alpha^3 + \left(\frac{1}{24} - \frac{2}{\pi^2} + \frac{16}{\pi^4} \right) \alpha^4 - i \left(\frac{143}{900\pi} - \frac{44}{9\pi^3} + \frac{32}{\pi^5} \right) \alpha^5 + \left(\frac{1}{720} - \frac{2171}{4050\pi^2} + \frac{104}{9\pi^4} - \frac{64}{\pi^6} \right) \alpha^6 + O(\alpha^7); \quad (70)$$

$$a_1 = -\frac{1}{2} \alpha^2 - \frac{i\alpha^3}{3\pi} + \left(\frac{1}{12} - \frac{2}{3\pi^2} \right) \alpha^4 - i \left(\frac{13}{90\pi} - \frac{4}{3\pi^3} \right) \alpha^5 + \left(\frac{1}{80} - \frac{49}{135\pi^2} + \frac{8}{3\pi^4} \right) \alpha^6 + O(\alpha^7); \quad (71)$$

$$a_2 = -\frac{1}{72} \alpha^4 - i \frac{\alpha^5}{60\pi} + \left(\frac{1}{240} - \frac{1}{30\pi^2} \right) \alpha^6 + O(\alpha^7); \quad (72)$$

$$a_3 = -\frac{\alpha^6}{3600} + O(\alpha^7). \quad (73)$$

The diffracted amplitude is

$$\frac{\pi}{2a} A_2(\theta) = \left\{ 1 + \left(\frac{1}{3} - \frac{4}{\pi^2} - \frac{1}{6} \sin^2 \theta \right) \alpha^2 + \left[\frac{16}{\pi^4} - \frac{20}{9\pi^2} + \frac{1}{15} + \left(\frac{2}{3\pi^2} - \frac{1}{15} \right) \sin^2 \theta + \frac{1}{120} \sin^4 \theta \right] \alpha^4 + \left[\left(\frac{112}{9\pi^4} + \frac{2}{315} - \frac{64}{\pi^6} - \frac{448}{675\pi^2} \right) + \left(\frac{16}{45\pi^2} - \frac{8}{3\pi^4} - \frac{1}{105} \right) \sin^2 \theta + \left(\frac{1}{280} - \frac{1}{30\pi^2} \right) \sin^4 \theta - \frac{\sin^6 \theta}{5040} \right] \alpha^6 \right\} + \frac{2}{\pi i} \left\{ \alpha + \left(-\frac{4}{\pi^2} + \frac{4}{9} - \frac{\sin^2 \theta}{6} \right) \alpha^3 + \left[\left(\frac{16}{\pi^4} - \frac{8}{3\pi^2} + \frac{71}{675} \right) + \left(\frac{2}{3\pi^2} - \frac{19}{270} \right) \sin^2 \theta + \frac{1}{120} \sin^4 \theta \right] \alpha^5 \right\} + O(\alpha^6). \quad (74)$$

The transmission coefficient is

$$t_2 = \frac{8}{\pi^2} \left[1 + \left(\frac{4}{9} - \frac{4}{\pi^2} \right) \alpha^2 + \left(\frac{71}{675} - \frac{8}{3\pi^2} + \frac{16}{\pi^4} \right) \alpha^4 + \left(\frac{568}{33075} - \frac{1936}{2025\pi^2} + \frac{128}{9\pi^4} - \frac{64}{\pi^6} \right) \alpha^6 \right] + O(\alpha^8). \quad (75)$$

Our expression for t_2 agrees completely with that obtained by C. J. Bouwkamp [1] and, where comparison is possible, our expressions for $\partial\phi_2/\partial z$ and $A_2(\theta)$ agree with the listed results in reference [1] (see p. 71).²¹ Note

²⁰ The maximum radius of convergence for this sum regarded as a function of α is not known. C. H. Yang [13] has shown that it is at least one-half.

²¹ A misprint occurs in the expression for $A_2(\theta)$ on page 71. It may be rectified by replacing $A_2(\theta)$ by $A_2(\theta)/4$.

however, that our calculations for $\partial\phi_2/\partial z$ and $A_2(\theta)$ have both been carried out to terms of order α^6 whereas the corresponding results listed in reference [1] have been carried out only as far as α^4 .

ACKNOWLEDGMENT

We would like to express our gratitude to Mrs. M. Vygantes, Miss L. Bader, Miss D. Pernova and Mr. H. Lehnson, members of the computing staff of the New York University, Institute of Mathematical Sciences, for checking our calculations and for proofreading the manuscript.

REFERENCES

- [1] C. J. Bouwkamp, "Diffraction theory," *Repts. Prog. in Phys.*, vol. 27, p. 35; 1954.
- [2] S. N. Karp, "The Natural Charge Distribution and Capacitance of a Finite Conical Shell, New York University, Inst. Math. Sci., New York, N. Y., EM Res. Div., Res. Rept. EM-35; 1956.
- [3] A. E. Green, and W. Zerna, "Theoretical Elasticity," Clarendon Press, Oxford, Eng.; 1954.
- [4] W. Magnus, "An Infinite System of Linear Equations Arising in Diffraction Theory," New York University, Inst. Math. Sci., New York, N. Y., EM Res. Div., Res. Rept. EM-80, 1951; and *Quart. Appl. Math.*, vol. 11, 1953.
- [5] N. Chako, "Characteristic Curves in Image Space," McGill Symp. on Microwave Optics, p. 79; April, 1959. Reported 1953.
- [6] D. S. Jones, "A New Method for Calculating Scattering with Particular Reference to the Circular Disc," New York University, Inst. Math. Sci., New York, N. Y., EM Res. Div. Res. Rept. EM-87, 1955; and, *Commun. Pure Appl. Math.*, vol. 9, p. 713; 1956.
- [7] N. I. Akhiezer and A. N. Akheizer, "K zadache o diffraktsii elektromagnitnykh voln u krugovogo otverstia v ploskom ek-rane," *Doklady Akad. Nauk S.S.S.R.*, vol. 109, p. 53; 1956.
- [8] B. Noble, "A Multiplying-Factor Method for the Solution of Certain Dual Integral Equations with Application to Diffraction by Strips and Discs," to be published.
- [9] R. C. MacCamy and A. E. Heins, "Axially Symmetric Solutions of Elliptic Differential Equations, Math. Dept., Carnegie Inst. Tech., Pittsburgh, Pa., Rept. 24; 1958.
- [10] W. Magnus and F. Oberhettinger, "Formulas and Theorems for the Functions of Mathematical Physics," Chelsea Press, New York, N. Y.; 1954.
- [11] N. I. Mushkelishvili, "Singular Integral Equations," P. Noordhoff, Groningen, Holland, 2nd ed., 1953.
- [12] J. Bazer and A. Brown, "Diffraction of Scalar Waves by a Circular Aperture," New York University, Inst. Math. Sci., New York, N. Y., EM Res. Div., Res. Rept. EM-44; 1959.
- [13] C. H. Yang, "On Some Fredholm Integral Equations Arising in Diffraction Theory," New York University Inst. Math. Sci., New York, N. Y., EM Res. Div., Res. Rept. BR-31, to be published.

Scalar Diffraction by an Elliptic Cylinder*

N. D. KAZARINOFF† AND R. K. RITT†

Summary—A recent method of the authors¹ is applied to the case of scalar scattering by a perfectly reflecting elliptic cylinder illuminated by waves from a line source parallel to the axis of the cylinder. The surface distribution in the shadow zone is calculated and the “creeping wave” representation for the scattered field in the shadow zone is derived. It is shown that the results are applicable if and only if $R_0\omega \gg 1$, where R_0 is the smallest radius of curvature on the cylinder and ω is the wave number.

IN a recent paper¹ we have developed a theory of scalar diffraction for bodies whose boundary surfaces are level surfaces in coordinate systems in which the scalar wave equation is separable. We applied the theory to the case of diffraction by a prolate spheroid and calculated the surface distribution. In this paper we make a similar application to the case of an elliptic cylinder, but, in addition, we consider the off-angle case and derive the “creeping wave” representation for the scattered field in the shadow zone. The surface distribution and the scattered field are described in Sections IV and V.

The asymptotic theory which we use to obtain our results gives us more terms of the series, in descending powers of ω , for the exponents of the “creeping wave” terms than does the elegant geometric theory of Keller.² These extra terms show that the condition $R_0\omega \gg 1$ is essential for our asymptotic theory to give a meaningful result for the diffracted field. It is reasonable to believe that the same restriction also applies to Keller’s geometric theory. More specifically, we show that it is the coefficient C of the attenuation exponent $CfR^{-2/3} ds$ which behaves in an unknown fashion when $R_0 \rightarrow 0$; the evaluations of C which have been made by Keller and the authors are performed under the hypothesis that $R_0\omega \gg 1$. In the Appendix, the magnitude of the attenuation term which Keller obtains is compared with the next term of the asymptotic series for the exponent in the case of two prolate spheroids. It is noted also that Levy³ has applied the geometric theory to the case of an elliptic cylinder and has given a mathematical derivation of the results thus obtained which is based upon the use of the Watson transform.

Generally, the analysis below closely follows that in our earlier paper which we shall henceforward refer to as

Reference 1.¹ Where this is true we only elucidate the principal points in the argument and omit most computations. Elsewhere a more complete discussion is given.

I. THE INTEGRAL REPRESENTATIONS

We consider an elliptic cylinder with semifocal distance c , eccentricity $e = \text{sech } \xi_0$, and semi-axes a and b . We introduce the (ξ, η, z) coordinate system defined by

$$\begin{aligned} x &= c \cosh \xi \cos \eta \\ y &= c \sinh \xi \sin \eta \\ z &= z. \end{aligned} \quad (1)$$

On the surface of the cylinder, $\xi = \xi_0$. For this cylinder, the operators L_ξ and L_η , appearing in Reference 1, equation (3.2), are defined by the formulas

$$-L_\xi u = \frac{d^2 u}{d\xi^2} + \gamma^2 \sinh^2 \xi u \quad (\xi \geq \xi_0)$$

and

$$-L_\eta u = \frac{d^2 u}{d\eta^2} + \gamma^2 \sin^2 \eta u \quad (-\pi \leq \eta \leq \pi).$$

The constant γ appearing above is $c(\omega - is)$, where s is a small positive number. The boundary conditions are

$$\left. \frac{du}{d\xi} \right|_{\xi_0} = 0 \quad \text{and} \quad u(\xi, \eta) = u(\xi, \eta + 2\pi). \quad (2)$$

Both the operators L_ξ and L_η are of the type considered in Reference 1, Section 3, and for them, respectively,

$$gq \equiv -g(\gamma^2 \sinh^2 \xi) \geq 2\omega sc^2 \sinh^2 \xi_0 > 0$$

and

$$gq \equiv -g(\gamma^2 \sin^2 \eta) \geq 0.$$

The radial operator L_ξ is to be considered on the interval $[\xi_0, \infty)$, $\xi_0 > 0$. For L_ξ , $p \equiv 1$, and hence $p(\xi_0) \neq 0$. In order to construct the resolvent Green’s function for L_ξ , we consider the homogeneous equation

$$L_\xi y - \lambda y = 0, \quad (3)$$

where

$$g\lambda < 2\omega sc^2 \sinh^2 \xi_0. \quad (4)$$

It has linearly independent solutions w_j , $j = 1, 2$, with the asymptotic forms

$$w_j(\xi) = (\sinh \xi)^{-1/2} e^{\pm i\gamma \cosh \xi} \left\{ 1 + \frac{\Theta(1)}{\xi} \right\},$$

in which $\Theta(1)$ denotes a function which is bounded for $\xi > N$, $|\gamma| > N$, and $|\lambda| < N$. In this and in succeeding

* The research reported in this paper has been sponsored by AF Cambridge Research Center, under Contracts AF 19(604)-1949 and AF 19(604)-4993.

† Dept. of Mathematics and Radiation Laboratory, Dept. of Elec. Engrg., University of Michigan, Ann Arbor, Mich.

¹ N. D. Kazarinoff and R. K. Ritt, “On the theory of scalar diffraction and its application to the prolate spheroid,” *Ann. Phys.*, vol. 6, pp. 277-299; March, 1959.

² J. B. Keller, “Diffraction by a convex cylinder,” *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. 4, pp. 312-321; July, 1956.

³ B. Levy, “Diffraction by an Elliptic Cylinder,” *Inst. Math. Sciences*, New York University, New York, N. Y., Rept. EM-121; December, 1958.

formulas, the upper sign is to be used when $j=1$, the lower one when $j=2$. Since $\Re(i\gamma) > 0$, the only solutions of (3) in $L^2(\xi_0, \infty)$ are multiples of w_2 ; therefore, L_ξ falls into Case I of Reference 1. We next single out the solution ϕ_1 of (3), which satisfies the boundary condition (2) and a solution ϕ_2 in $L^2(\xi_0, \infty)$:

$$\phi_1(\xi, \lambda) = w_1(\xi, \lambda)w_2'(\xi_0, \lambda) - w_2(\xi, \lambda)w_1'(\xi_0, \lambda)$$

$$\phi_2(\xi, \lambda) = w_2(\xi, \lambda).$$

At ξ_0 , $\phi_1 = -2i\gamma$. Thus, the resolvent Green's function is

$$G(\xi, \xi', \lambda) = \frac{-1}{2i\gamma w_2'(\xi_0, \lambda)} \begin{cases} \phi_1(\xi) \phi_2(\xi') & (\xi < \xi') \\ \phi_2(\xi) \phi_1(\xi') & (\xi' < \xi) \end{cases}.$$

$$v(\xi, \eta, \Xi, \tau) = \frac{1}{4\pi\gamma} \int_{\Gamma} \frac{[w_1(\xi, \lambda)w_2'(\xi_0, \lambda) - w_2(\xi, \lambda)w_1'(\xi_0, \lambda)]w_2(\Xi, \lambda)}{w_2'(\xi_0, \lambda)} \tilde{G}(\eta, \tau, -\lambda) d\lambda. \quad (8)$$

The operator L_η is to be considered on $[-\pi, \pi]$. To construct its resolvent Green's function, we consider the homogeneous equation

$$L_\eta y - (-\lambda y) = 0,$$

where λ satisfies the condition (4). Using the notation of Meixner and Schäfke,⁴ let $y_I(\eta, -\lambda)$ and $y_{II}(\eta, -\lambda)$ be the solutions of the homogeneous equation for which

$$y_I(0) = 1, \quad y_I'(0) = 0,$$

$$y_{II}(0) = 0, \quad y_{II}'(0) = 1.$$

It is then a routine computation to show that the resolvent Green's function for the periodic problem is

$$\tilde{G}(\eta, \tau, -\lambda) = \frac{1}{2} \left\{ \frac{y_I(\tau)y_I(\pi - \eta)}{y_I'(\pi)} + \frac{y_{II}(\tau)y_{II}(\pi - \eta)}{y_{II}'(\pi)} \right\} \quad (5)$$

for $\tau < \eta$. The relation

$$\tilde{G}(\tau, \eta, -\lambda) = \tilde{G}(\eta, \tau, -\lambda) \quad (6)$$

then serves to define \tilde{G} for $\eta < \tau$.

We are now in a position to write down the contour integral representation, guaranteed by the theory in Reference 1, for the solution $v(\xi, \eta)$ of

$$[\nabla^2 + (\omega - is)^2]v = \rho(\Xi, \tau) \quad (s > 0)$$

which we seek. The function $\rho(\Xi, \tau)$ is a distribution corresponding to a line source at (Ξ, τ) . The representation is

$$v(\xi, \eta, \Xi, \tau) = \frac{1}{2\pi i} \int_{\Gamma} \tilde{G}(\eta, \tau, -\lambda) G(\xi, \Xi, \lambda) d\lambda,$$

where Γ is a path in the λ -plane defined by the conditions

$$\lambda = l + i\delta, \quad 0 < \delta < 2\omega s c^2 \sinh^2 \xi_0,$$

in which l and δ are real. The integration path Γ is oriented in the direction of increasing l . When $s \rightarrow 0^+$, $v(\xi, \eta, \Xi, \tau)$ reduces to the Green's function for the elliptic cylinder relative to the line source at (Ξ, τ) .

The distribution on the surface of the cylinder and off the surface will be considered separately. On the cylinder,

$$v(\xi_0, \eta, \Xi, \tau) = \frac{1}{2\pi i} \int_{\Gamma} \frac{w_2(\Xi, \lambda) \tilde{G}(\eta, \tau, -\lambda)}{w_2'(\xi_0, \lambda)} d\lambda. \quad (7)$$

As in Reference 1, we shall evaluate this integral by residues, the residues contributed by the zeros of $w_2'(\xi_0, \lambda)$. If $\Xi > \xi > \xi_0$,

In the case where it is practical to evaluate this integral by the residues contributed by the zeros of $w_2'(\xi_0, \lambda)$, the representation (8) reduces to

$$v(\xi, \eta, \Xi, \tau) = \frac{-1}{4\pi\gamma} \int_{\Gamma} \frac{w_2(\xi, \lambda)w_1'(\xi_0, \lambda)w_2(\Xi, \lambda)}{w_2'(\xi_0, \lambda)} \tilde{G}(\eta, \tau, -\lambda) d\lambda. \quad (9)$$

II. THE TURNING POINT ANALYSIS

The first objective is to determine the zeros of $w_2'(\xi_0, \lambda)$. We need only sketch the analysis in view of its similarity to that in Reference 1, Sections 6 and 7. The differential equation satisfied by the w_j is

$$\frac{d^2 y}{d\xi^2} + (\gamma^2 \sinh^2 \xi + \lambda)y = 0.$$

If we let

$$\lambda = -\gamma^2 \sinh^2 \xi_1, \quad (10)$$

then this takes the form

$$\frac{d^2 y}{d\xi^2} + \gamma^2 (\sinh^2 \xi - \sinh^2 \xi_1)y = 0.$$

We define

$$\phi^2(\xi, \xi_1) = \sinh^2 \xi - \sinh^2 \xi_1,$$

$$\Phi(\xi, \xi_1) = \int_{\xi_1}^{\xi} \phi(t, \xi_1) dt, \quad \zeta(\xi, \xi_1, \gamma) = \gamma \Phi(\xi, \xi_1),$$

and

$$\Psi(\xi, \xi_1) = \Phi^{1/6}(\xi, \xi_1) \phi^{-1/2}(\xi, \xi_1) \quad (\xi \neq \xi_1),$$

with

$$\Psi(\xi_1, \xi_1) = \lim_{\xi \rightarrow \xi_1} \Psi(\xi, \xi_1).$$

In terms of the above notation, the solutions w_j have the asymptotic forms

⁴ J. Meixner and F. Schäfke, "Mathieusche Funktionen und Spheroidfunktionen," Springer-Verlag, Berlin, Ger., pp. 98-100; 1954.

$$w_j = \gamma^{1/6} e^{\mp i\gamma f(\xi_1)} \left\{ V^{(j)}(\xi) + \frac{B(\xi, \gamma)}{\gamma} \right\}$$

$$w_j' = \gamma^{1/6} e^{\mp i\gamma f(\xi_1)} \{ V^{(j)'}(\xi) + B(\xi, \gamma) \}$$

when $|\zeta| < N$, and the forms

$$w_j = \phi^{-1/2}(\xi, \xi_1) e^{\pm i\gamma[f(\xi) - f(\xi_1)]} [1 + B(\zeta^{-1})]$$

when $|\zeta| > N$. In these formulas, B is used generically for a function which is uniformly bounded for the range of ζ in question and for $|\gamma| > N$,

$$f(\xi_1) = - \int_0^{\pi/2} \sqrt{\sinh^2 \xi_1 + \sin^2 \theta} d\theta,$$

and

$$V^{(j)}(\xi) = \left(\frac{\pi}{2} \right)^{1/2} e^{\pm 5\pi i/12} \Psi(\xi) \zeta^{1/3} H_{1/3}^{(j)}(\zeta),$$

where $H_{1/3}^{(j)}$ is a Hankel function.

When $|\lambda| \ll |\gamma|^2$, it can be seen from these formulas that $w_2'(\xi_0, \lambda)$ has no zeros. Provided (ξ_0, η) is not too close to the shadow boundary, the zeros of $w_2'(\xi_0, \lambda)$ corresponding to values of λ with $|\lambda| \gg |\gamma|^2$ have large imaginary parts, and the terms which they contribute to the residue series may be neglected (see, for example, Franz⁵ and Levy³). When $|\lambda|$ is comparable to $|\gamma|^2$, $w_2'(\xi, \lambda)$ vanishes only if

$$\frac{d}{d\zeta} \{ \zeta^{1/3} H_{1/3}^{(2)}(\zeta) \} + \mathcal{O}(\gamma^{-1}) = 0.$$

If $\xi = \xi_0$, ξ_0 fixed, and λ is considered as variable, the value ζ_r of ζ which corresponds to the r th zero of $w_2'(\xi_0, \lambda)$ may be thought of as the value of

$$\gamma \int_{\xi_1(\lambda, \gamma)}^{\xi_0} \phi(t, \xi_1(\lambda, \gamma)) dt,$$

which is attained when $\lambda = \lambda_r$, since ξ_1 and λ are related by (10). Thus

$$\zeta_r = \zeta(\xi_0, \xi_1(\lambda_r, \gamma)).$$

Because the zeros h_r of

$$\frac{d}{dt} [t^{1/3} H_{1/3}^{(2)}(t)]$$

are simple and because this function is analytic in a neighborhood of each of its zeros,

$$\zeta_r = h_r + \mathcal{O}(\gamma^{-1}).$$

The zeros h_r and the values of related functions such as $t H_{1/3}^{(2)}(t)$ at these zeros are known.⁶

The relation

$$\int_{\xi_1(\lambda_r, \gamma)}^{\xi_0} \phi(t, \xi_1(\lambda_r, \gamma)) dt = h_r \gamma^{-1} + \mathcal{O}(\gamma^{-2})$$

⁵ W. Franz, "Ueber die Greenschen Funktionen des Zylinders und der Kugel," *Z. Naturforsch.*, vol. 9a, pp. 705-716; September, 1954.

⁶ British Association Mathematical Tables, "The Airy Integral," Cambridge University Press, London, Eng., and New York, N. Y.; 1946.

may now be used to compute $\xi_r = \xi_1(\lambda_r, \gamma)$ by expanding the integrand on the left-hand side in powers of $(t - \xi_1)$ or $(t - \xi_0)$. It is vital to note that both of these expansions are slowly convergent as $\xi_0 \rightarrow 0$. Therefore, the approximation for ξ_r which we obtain by neglecting all but the first two terms is not useful when $\xi_0 \rightarrow 0$ and ω is fixed. Henceforth, we assume that ξ_0 is bounded away from zero. We find, under this hypothesis, that

$$\begin{aligned} \xi_r - \xi_0 &= \frac{e^{-\pi i/3} 3^{2/3}}{2(\sinh \xi_0 \cosh \xi_0)^{1/3}} \left(\frac{h_r}{\gamma} \right)^{2/3} \\ &\cdot \left\{ 1 - \frac{7e^{-\pi i/3} 3^{2/3} (\sinh^2 \xi_0 + \cosh^2 \xi_0)}{60(\sinh \xi_0 \cosh \xi_0)^{4/3}} \left(\frac{h_r}{\gamma} \right)^{2/3} \right\} \\ &+ \mathcal{O}(\gamma^{-5/3}). \end{aligned} \quad (11)$$

The specific value of λ_r will not be needed.

Computation of

$$\frac{\partial w_2'(\xi_0, \lambda)}{\partial \xi_1} \frac{\partial \xi_1}{\partial \lambda} \Big|_{\lambda_r}$$

now leads to an approximation for the residue contribution of w_2' at λ_r .

$$\begin{aligned} \frac{\partial}{\partial \lambda} w_2'(\xi_0, \lambda) \Big|_{\lambda_r} &= \left\{ \frac{3\pi}{8\gamma \sinh \xi_0 \cosh \xi_0} \right\}^{1/2} e^{i\gamma f(\xi_r) + 5\pi i/4} h_r H_{1/3}^{(2)}(h_r) \\ &\cdot [1 + \mathcal{O}(\gamma^{-1/3})]. \end{aligned} \quad (12)$$

III. THE ANGULAR GREEN'S FUNCTION

In subsequent work, it will be necessary also to have an approximation for $w_1'(\xi_0, \lambda_r)$. An easy calculation yields the formula

$$\begin{aligned} w_1'(\xi_0, \lambda_r) &= \left(\frac{\pi}{2} \right)^{1/2} (3\gamma^5 \sinh \xi_0 \cosh \xi_0)^{1/6} e^{-i\gamma f(\xi_r) + 3\pi i/4} \\ &\cdot h_r^{1/3} [\zeta^{1/3} H_{1/3}^{(1)}(\zeta)]'_{\zeta=h_r} \{ 1 + \mathcal{O}(\gamma^{-2/3}) \}. \end{aligned}$$

Next, $\tilde{G}(\eta, \tau - \lambda_r)$ is approximated. For $\lambda_r = -\gamma^2 \sinh^2 \xi_r$, the Liouville asymptotic representations for y_I and y_{II} are

$$y_I \sim \left[\frac{K_r(0)}{K_r(\eta)} \right]^{1/2} \cos \left[\gamma \int_0^\eta K_r(t) dt \right] \quad (13)$$

and

$$y_{II} \sim \frac{\sin \left(\gamma \int_0^\eta K_r(t) dt \right)}{\gamma [K_r(0) K_r(\eta)]^{1/2}},$$

where

$$K_r(t) = [\sin^2 t + \sinh^2 \xi_r]^{1/2}. \quad (14)$$

These results and the relations (5), (6), and (13) combine to yield the formula

$\tilde{G}(\eta, \tau, -\lambda)$

$$\sim \frac{-\cos\left(\gamma\left[\int_0^\pi K_r(t)dt - \left|\int_\tau^\eta K_r(t)dt\right|\right]\right)}{2\gamma \sin\left(\gamma\int_0^\pi K_r(t)dt\right)[K_r(\tau)K_r(\eta)]^{1/2}}. \quad (15)$$

IV. THE SURFACE DISTRIBUTION

In this section the surface distribution is considered and the residue series for the integral in (7) is derived. The residue series is a sum of terms of the form

$$\frac{w_2(\Xi, \lambda_r)\tilde{G}(\eta, \tau, -\lambda_r)}{\frac{\partial}{\partial \lambda}[w_2'(\xi_0, \lambda)]_{\lambda=\lambda_r}}.$$

$$\frac{i\left\{\exp\left[-i\gamma\left\{\left|\int_\tau^\eta K_r(t)dt\right| - \int_0^{\pi/2} K_r(t)dt\right\}\right] + \exp\left[-i\gamma\left\{3\int_0^{\pi/2} K_r(t)dt - \left|\int_\tau^\eta K_r(t)dt\right|\right\}\right]\right\}}{1 - \exp\left(-4i\gamma\int_0^{\pi/2} K_r(t)dt\right)}. \quad (19)$$

These can be approximated by using the results obtained above. We make use of such approximations in order to investigate the convergence of the residue series and to see if the boundary of the region of convergence coincides with the geometric shadow boundary. Further approximations can be made by expanding in powers of $(\xi_r - \xi_0)$ and neglecting the terms which cannot be specifically computed using the estimate (11) for $\xi_r - \xi_0$. Henceforward, we shall assume that the parameter s involved in γ is zero.

First consider $w_2(\Xi, \lambda_r)$. Since Ξ is large,

$$w_2(\Xi, \lambda_r) = \exp\left\{-i\gamma\left[\int_{\xi_r}^\Xi H_r(t)dt + \int_0^{\pi/2} K_r(\theta)d\theta\right]\right\} \frac{[1 + O(\gamma^{-1}e^{-\Xi})]}{[H_r(\Xi)]^{1/2}},$$

where

$$H_r(t) = (\sinh^2 t - \sinh^2 \xi_r)^{1/2}. \quad (16)$$

$$v(\xi_0, \eta, \Xi, \tau) \sim \frac{e^{2\pi i/4}}{2\gamma} \left\{ \frac{2\gamma \sinh \xi_0 \cosh \xi_0}{3\pi} \right\}^{1/2}.$$

$$\cdot \sum_r \frac{\exp\left\{-i\gamma\left[(\coth \Xi)H_r(\Xi) + \int_0^{l_r(\Xi)} K_r(t)dt + \left|\int_\tau^\eta K_r(t)dt\right| - \int_0^{\pi/2} K_r(t)dt\right]\right\}}{h_r H_{1/3}^{(2)}(h_r)[K_r(\eta)K_r(\tau)H_r(\Xi)]^{1/2}}, \quad (21)$$

Performing an integration by parts, we find that

$$w_2(\Xi, \lambda_r) = \exp\left\{-i\gamma\left[(\coth \Xi)H_r(\Xi) + \int_0^{l_r(\Xi)} K_r(t)dt\right]\right\} \frac{[1 + O(\gamma^{-1}e^{-\Xi})]}{[H_r(\Xi)]^{1/2}}, \quad (17)$$

with

$$l_r(t) = \sin^{-1} \left\{ \frac{\sinh \xi_r}{\sinh t} \right\}. \quad (18)$$

The exponential factors in the remainder of the r th residue term are

$$\left[\exp\left(-i\gamma\int_0^{2\pi} K_r(t)dt\right)\right] \frac{\cos\left[\gamma\left\{\int_0^\pi K_r(t)dt - \left|\int_\tau^\eta K_r(t)dt\right|\right\}\right]}{\sin\left[\gamma\int_0^\pi K_r(t)dt\right]},$$

or

It is important to note that because

$$\int_0^{\pi/2} K_r(t)dt = \int_0^{\pi/2} K_0(t)dt + (\xi_r - \xi_0) \int_0^{\pi/2} \frac{\sinh \xi_0 \cosh \xi_0 dt}{K_0(t)} + \dots,$$

and because

$$\Im(\xi_r - \xi_0) < 0, \quad \Re\left(i\gamma\int_0^{\pi/2} K_r(t)dt\right) > 0. \quad (20)$$

Since $\Re(\gamma)$ is large, it follows that the dominant term in (19) is

$$\exp\left\{-i\gamma\left[\left|\int_\tau^\eta K_r(t)dt\right| - \int_0^{\pi/2} K_r(t)dt\right]\right\}.$$

Therefore, by the relations (7), (17), (15), (12), and (19), the surface distribution

where only the dominant part of the first creeping wave term has been included. In this formula the functions H_r , K_r and l_r are defined by the relations (16), (14), and (18), respectively.

Let us assume that $\eta > \tau$. Then if we expand K_r in powers of $(\xi_r - \xi_0)$, we find that the exponential terms in the above summation take the form

$$\exp(-i\gamma) \left\{ \left(\int_{\tau}^{\eta} + \int_0^{i_0(\Xi)} - \int_0^{\pi/2} \right) K_0(t) dt + (\coth \Xi) H_0(\Xi) \right. \\ \left. + (\sinh \xi_0 \cosh \xi_0) (\xi_r - \xi_0) \left(\int_{\tau}^{\eta} + \int_0^{i_0(\Xi)} - \int_0^{\pi/2} \right) \frac{dl}{K_0(l)} + \frac{(\xi_r - \xi_0)^2}{2} \left(\int_{\tau}^{\eta} + \int_0^{i_0(\Xi)} - \int_0^{\pi/2} \right) \right. \\ \left. \cdot \left(\frac{(\sinh^4 \xi_0 + (\cosh^2 \xi_0 + \sinh^2 \xi_0) \sin^2 t)}{K_0^3(t)} + \frac{\tanh \Xi \cosh \xi_0}{H_0(\Xi)} \right) dt \right\}.$$

Each of the above integrals is a real quantity. The only terms with an imaginary part in the above exponent are $-i\gamma$, $(\xi_r - \xi_0)$ and $(\xi_r - \xi_0)^2$. Using the transformation (1), we find that $c \int K_0(t) dt$ represents an integral $\int ds$ where ds is the differential of arc on the ellipse $b^2 x^2 + a^2 y^2 = a^2 b^2$ and that $(ab)^{2/3} \cdot c^{-1} \cdot \int K_0^{-1}(t) dt$ represents an integral $\int R^{-2/3} ds$, where R is the local radius of curvature on the ellipse. This tells us that to a first approximation, the attenuation of the "creeping waves" is as predicted by Keller.² Now by (11),

$$\gamma(\sinh \xi_0 \cosh \xi_0) (\xi_r - \xi_0) \sim \frac{(ab)^{2/3}}{c} \cdot C_1,$$

where

$$C_1 = \frac{(-9\omega h_r^2)^{1/3}}{2} \left[1 - \frac{7e^{-\pi i/3}}{60} \left(\frac{3h_r}{\omega} \right)^{2/3} \left(R_0^{-2/3} + \frac{R_0^{1/3}}{a} \right) \right].$$

$R_0 = b^2/a$ is the radius of curvature at the ends of the major axis of the ellipse. Thus the second term in the above exponent becomes

$$\alpha(\tau, \eta) = \frac{\cosh \xi_0 \{ \cos \tau \sin \eta (\cosh^2 \xi_0 - \cos^2 \eta)^{1/2} - \cos \eta \sin \tau (\cosh^2 \xi_0 - \cosh^2 \tau)^{1/2} \}}{\cosh^2 \xi_0 - \cos^2 \tau \cos^2 \eta}.$$

$$+ iC_1 \left(\int_{s(\tau)}^{s(\eta)} + \int_{s(0)}^{s(i_0(\Xi))} - \int_{s(0)}^{s(\pi/2)} \right) R^{-2/3} ds.$$

The formula for C_1 reveals an essential limitation upon our theory and perhaps that of Keller; namely, the above expansion of the creeping wave exponents in descending powers of ω is meaningful only if $R_0 \omega \gg 1$. In particular, for a fixed ω , R_0 cannot be taken too small. Thus, we have derived mathematically the expected physical restriction upon theories of this kind.

The $(\xi_r - \xi_0)^2$ term in the exponent is of less interest. We have carried out its computation only in the case of the prolate spheroid; see the Appendix for the results. As the discussion there would indicate, when $\xi_0 \rightarrow \infty$ and the cylinder becomes circular the $(\xi_r - \xi_0)$ and $(\xi_r - \xi_0)^2$ terms cancel in such a way as to produce the expected exponent.

It remains to investigate the convergence of the residue series when summed in the "creeping wave" form. The condition (20) shows that the convergence will not be rapid unless

$$f(\xi, \eta) \equiv \left(\int_{\tau}^{\eta} + \int_0^{i_0(\Xi)} - \int_0^{\pi/2} \right) K_0^{-1}(t) dt > 0.$$

In fact, for a given τ , the condition

$$f(\tau, \eta) = 0 \quad (22)$$

determines the boundary of the region of convergence. We shall show that this is indeed the optical shadow boundary. To do this, it is convenient to put the elliptic integrals of the first kind involved in $f(\xi, \eta)$ into Legendre form. We then find

$$f(\xi, \eta) = \left(\int_0^{\cos \tau} + \int_0^{-\cos \eta} - \int_0^{\sqrt{1 - \frac{\sinh^2 \xi_0}{\sinh^2 \Xi}}} \right) \frac{edl}{[(1-t^2)(1-e^2 t^2)]^{1/2}}.$$

The addition formula for integrals of the first kind can now be used to write

$$\int_0^{\cos \tau} + \int_0^{-\cos \eta} \text{ as } \int_0^{\alpha},$$

where

A tangent to the ξ_0 -ellipse drawn from the point (Ξ, τ) touches the ellipse at the point (ξ_0, η) such that

$$\cosh \Xi \sinh \xi_0 \cos \tau \cos \eta + \sinh \Xi \cosh \xi_0 \sin \tau \sin \eta = \sinh \xi_0 \cosh \xi_0. \quad (23)$$

When $\tau = 0$, it is a trivial matter to verify that (22) and (23) are equivalent. For nonzero τ , the verification is easy but tedious.

V. THE FAR FIELD

Lastly, we discuss the far field in the shadow zone. In particular, we derive the residue series for the integral in (9) for large ξ and $\Xi > \xi > \xi_0$, or for large Ξ and $\xi > \Xi > \xi_0$. The r th residue R_r is precisely

$$(-2i\gamma)^{-1} w_2(\xi, \lambda_r) w_1'(\xi_0, \lambda_r)$$

times the r th residue in the case of the surface distribution. It therefore follows from the relations (17), (13), and (21) that

$$R_r = F_r E_r,$$

where

$$R_r = \frac{(\sinh \xi_0 \cosh \xi_0)^{2/3} [\xi^{1/3} H_{1/3}^{(1)}(\xi)]'_{\xi=h_r}}{2 \cdot 3^{1/3} (\gamma h_r)^{2/3} H_{1/3}^{(2)}(h_r) [H_r(\Xi) H_r(\xi) K_r(\eta) K_r(\tau)]^{1/2}}$$

and

$$E_r = \frac{\cos \left\{ \gamma \left[\int_0^\pi K_r(t) dt - \left| \int_\tau^\eta K_r(t) dt \right| \right] \right\}}{\sin \left(\gamma \int_0^\pi K_r(t) dt \right)} \cdot \exp \left\{ -i\gamma \left[(\coth \xi) H_r(\xi) + (\coth \Xi) H_r(\Xi) + \left(\int_0^{l_r(\xi)} + \int_0^{l_r(\Xi)} - \int_0^\pi \right) K_r(t) dt \right] \right\}.$$

In these formulas the functions H_r , K_r , and l_r are defined by the relations (16), (14), and (18), respectively.

We can also write E_r as a sum of creeping wave terms

$$E_r = i \sum_{n=0}^{\infty} \left\{ \exp \left[i\gamma \left(\left| \int_\tau^\eta \right| - (2n-1) \int_0^\pi \right) K_r(t) dt \right] + \exp \left[-i\gamma \left((2n+1) \int_0^\pi - \left| \int_\tau^\eta \right| \right) K_r(t) dt \right] \right\} \cdot \exp \left\{ -i\gamma \left[(\coth \xi) H_r(\xi) (\coth \Xi) H_r(\Xi) + \left(\int_0^{l_r(\xi)} + \int_0^{l_r(\Xi)} \right) K_r(t) dt \right] \right\}.$$

Let

$$I_r(f) = -i\gamma \left[\left| \int_\tau^\eta \right| - (2n-1) \left(\int_0^\pi + \int_0^{l_r(\xi)} + \int_0^{l_r(\Xi)} \right) \right] f(t) dt.$$

Then a typical exponent is

$$-i\gamma (\coth \xi) H_0(\Xi) - i\gamma (\coth \Xi) H_0(\xi) - i\gamma I_0(K_0) + iC_1 \frac{(ab)^{2/3}}{c} I_0(K_0^{-1}) + \dots$$

The description of the terms involving I_0 in terms of physical parameters is essentially the same as that given at the end of Section IV; and, of course, the remarks made there upon the region of validity of the expansion also apply. It is a considerably more tedious matter to verify that the creeping wave expansion converges in the geometric shadow zone.

APPENDIX

In Reference 1 we developed an expression for the surface distribution induced by a plane wave whose plane is perpendicular to the axis of the spheroid. This expression is in the form of the well-known "creeping wave" representation, namely

$$\sum_{n=0}^{\infty} (-1)^n \sum_r A_r \left\{ \exp \left(i\nu_r [d_r(\eta) + nL_r] + \frac{\pi i}{4} \right) + \exp \left(i\nu_r [d_r^*(\eta) + nL_r] - \frac{\pi i}{4} \right) \right\}.$$

We should like to point out the restriction upon our result and that given by the geometrical optics theory of Keller.² The restriction stems from the terms $(d_r(\eta) + nL_r)$ and $(d_r^*(\eta) + nL_r)$. We have shown that

$$i\nu_r d_r(\eta) = i\sqrt{a^2 - b^2} \omega \int_0^\eta \left(\frac{\xi_r^2 - t^2}{1 - t^2} \right)^{1/2} dt,$$

where a and b are the semimajor and minor axes of the spheroid, ω is the wave number, and ξ_r is related to the r th zero of the Airy function (see Section 10 of Reference 1). If one expands the integral above in a series of ascending powers of $\xi_r - \xi_0$, where $\xi_0^{-1} = e$ is the eccentricity, one obtains the result (taking into account only the first 3 terms of the expansion)

$$i\nu_r d_r(\eta) \sim i\omega \int_{\text{Arc cos } \eta}^{\pi/2} ds + i(c_1 \omega^{1/3} + c_2 \omega^{-1/3}) \int_{\text{Arc cos } \eta}^{\pi/2} R^{-2/3} ds + i c_3 \omega^{-1/3} \int_0^\eta \frac{dt}{[(1 - e^2 t^2)^3 (1 - t^2)]^{1/2}},$$

where R is the local radius of curvature, s is arc length,

$$c_1 = \frac{e^{-\pi i/3} (3h_r)^{2/3}}{2}$$

$$c_2 = \frac{e^{-2\pi i/3} (3h_r)^{4/3}}{8} \left\{ \frac{(8 - 7R_0/a)}{15R_0^{2/3}} + \frac{R_0^{1/3}}{a} \right\}$$

$$c_3 = - \frac{e^{-2\pi i/3} (3h_r)^{4/3} R_0^{2/3}}{8a}.$$

$R_0 = b^2/a$ is the radius of curvature at the tip of the spheroid, and h_r is the r th zero of $[t^{1/3} H_{1/3}^{(2)}(t)]'$.

In the geometrical optics theory of Keller² only the terms $\int ds$ and $c_1 \int R^{-2/3} ds$ are present.

Two observations of interest can be made from these formulas. Firstly, we note that in the case of the sphere ($e=0$ and $b=a$) the term

$$\frac{i e^{-2\pi i/3} (3h_r)^{4/3} R_0^{1/3}}{\omega^{1/3} a} \int R^{-2/3} ds$$

is the negative of the term

$$i c_3 \omega^{-1/3} \int_0^\eta \frac{dt}{[(1 - e^2 t^2)^3 (1 - t^2)]^{1/2}}.$$

This is consistent with the known results for the sphere. Secondly, we observe that if $R_0 \rightarrow 0$, that is $e \rightarrow 1$, and if ω is fixed, the c_2 and c_3 terms completely dominate the attenuation. Let us estimate $g c_2 / (\omega^{2/3} g c_1)$ in the case $n=r=0$, the case of the most significant term in the creeping wave representation. One finds that

$$\frac{g c_2}{\omega^{2/3} g c_1} = \frac{|3h_0|^{2/3}}{\omega^{2/3} 4} \left\{ \frac{(8 - 7R_0/a)}{15R_0^{2/3}} + \frac{R_0^{1/3}}{a} \right\},$$

where $|h_0|^{2/3} \approx \frac{3}{2} (1.0188)$.

If $a = 6$ inches, $b = 0.6$ inch, and $\lambda = 1.25$ inches,

$$\frac{g_{c_2}}{\omega^{2/3} g_{c_1}} \approx 3/8;$$

while if $a = 1$ inch, $b = 0.1$ inch, and $\lambda = 1.25$ inches,

$$\frac{g_{c_2}}{\omega^{2/3} g_{c_1}} \approx 4.$$

Therefore, for such spheroids and such a λ it appears that our theory and that of Keller will not give a significant result for the diffracted field. That is to say, the condition $R_0\omega \gg 1$ is essential for the expansion of the creeping wave exponents in descending powers of ω to be meaningful. This condition is a consequence of the fact that in order for the expansion of

$$\int_0^\pi \left(\frac{\xi_r^2 - t^2}{1 - t^2} \right)^{1/2} dt$$

in powers of $\xi_r - \xi_0$ to be useful, we must know $\xi_r - \xi_0$; whereas $\xi_r - \xi_0$ has been estimated under the hypothesis that ξ_0 is bounded away from 1. Further, even if $R_0\omega \gg 1$, since our theory is only an asymptotic one, we have no *a priori* way of predicting for a particular choice of parameters whether or not the additional terms in this series, which we have found, give a more or less accurate result than that obtained by consideration of only the first term. The examples $a = 6$ inches, 1 inch, have been chosen for discussion because of recent measurements of Olte and Silver⁷ on such spheroids.

⁷ A. Olte and S. Silver, "New results on backscattering from cones and spheroids," URSI-Toronto Symposium; June, 1959.

Fock Theory—An Appraisal and Exposition*

R. F. GOODRICH†

Summary—We present an exposition and certain generalizations of recent work on a class of problems in classical electromagnetic theory. Briefly, we indicate the approach as the Fock theory which is a method of obtaining the field induced by an incident electromagnetic wave on, or near, the surface of a good conductor. The surface is restricted to be smooth, convex, and of characteristic dimensions which are "large" with respect to the wavelength of the incident radiation.

TWO-DIMENSIONAL PROBLEMS

AT the outset we point out that Fock's method is an essentially two-dimensional one which has as its prototype the solution of the diffraction of electromagnetic radiation by a perfectly conducting, infinite circular cylinder. For this reason we propose to review the solution of the circular cylinder problem, following, and somewhat adding to, the treatment of W. Franz.¹ We consider a perfectly conducting circular cylinder of radius a , having its axis along the Z axis of a Cartesian coordinate system. Let a plane wave be incident along the X axis. (See Fig. 1.)

We now wish to determine the magnetic field induced on the surface of the cylinder. If the incident magnetic field is in the Z direction

$$H_0 = e^{ikz}\hat{Z}, \quad (1)$$

the only nonvanishing component of the field on the surface is in the Z direction. Hence, we write

$$H = \hat{Z}\psi. \quad (2)$$

The function ψ is then required to satisfy the equation

$$(\nabla^2 + k^2)\psi = 0, \quad (3)$$

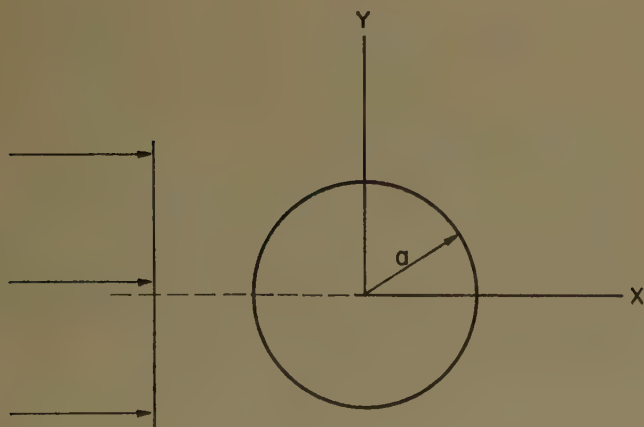


Fig. 1.

where $k = 2\pi/\lambda$, λ the wavelength, and the boundary condition

$$\left. \frac{\partial \psi}{\partial r} \right|_{r=a} = 0. \quad (4)$$

Using a method of R. K. Ritt and N. D. Kazarinoff² we can immediately write the solution as

$$\psi(a, \phi) = \frac{1}{\pi k a} \int_{-\infty + i\epsilon}^{\infty + i\epsilon} d\nu \frac{e^{i\nu(\phi - 3\pi/2)} + e^{-i\nu(\phi - \pi/2)}}{\sin \nu \pi H_{\nu}^{(1)'}(ka)}, \quad (5)$$

where ϕ is the polar angle. But since $\text{Im } \nu > 0$, we can make the convergent expansion

$$\frac{1}{\sin \nu \pi} = -2ie^{i\nu\pi} \sum_{m=0}^{\infty} e^{2i\nu m\pi}. \quad (6)$$

Substituting in (5),

$$\psi = \frac{2i}{\pi k a} \sum_{m=0}^{\infty} \int \frac{d\nu}{H_{\nu}^{(1)'}(ka)} [e^{i\nu(\theta + 2\pi m)} + e^{i\nu(\theta' + 2\pi m)}], \quad (7)$$

where we have put

$$\begin{aligned} \theta &= \phi - \frac{\pi}{2}, \\ \theta' &= \frac{3\pi}{2} - \phi. \end{aligned} \quad (8)$$

So, quite generally, we have the problem of solving integrals of the form

$$I(\theta) = \int_{-\infty}^{\infty} d\nu \frac{e^{i\nu\theta}}{H_{\nu}^{(1)'}(ka)}, \quad (9)$$

where θ lies in the range $[-\pi/2, \infty]$.

We assume $ka \gg 1$ and propose the use of certain asymptotic approximations to the Hankel function. The appropriate asymptotic forms are found by an examination of the stationary points of the integrand.

Using Langer's uniform asymptotic form of the Hankel function³

$$\begin{aligned} H_{\nu}^{(1)}(ka) \\ = e^{i\pi/6} \sqrt{\frac{\sin \alpha - \cos \alpha}{\sin \alpha}} H_{1/3}^{(1)} ka (\sin \alpha - \cos \alpha) \end{aligned} \quad (10)$$

* The research reported in this paper has been sponsored by the AF Cambridge Research Center, under Contract AF 19(604), 1949.

† Radiation Laboratory, University of Michigan, Ann Arbor, Mich.

¹ W. Franz, "Ueber die Greenschen Funktionen des Zylinders und der Kugel," *Z. Naturforsch.*, vol. 9a, pp. 705-716; 1954.

² R. K. Ritt and N. D. Kazarinoff, "On the theory of scalar diffraction and its application to the prolate spheroid," *Ann. Phys.*, vol. 6, pp. 277-299; March, 1959.

³ Staff of the Bateman Manuscript Project, "Higher Transcendental Functions," McGraw-Hill Book Co., Inc., New York, N. Y., vol. II; 1955.

where $\nu = ka \cos \alpha$. We find that the integrand has the phase

$$\phi = \nu\theta - ka(\sin \alpha - \cos \alpha). \quad (11)$$

Hence, the phase is stationary at $\alpha = -\theta$ or

$$\nu = ka \cos \theta. \quad (12)$$

We now draw a distinction between the regions $\theta \sim -\pi/2$ and θ near zero or positive. The first region corresponds to the physical region of direct illumination, *i.e.*, the geometrical optics region, and, by carrying out a stationary phase⁴ evaluation of the integral, we find

$$I(\theta) \cong \pi ka e^{ika \sin \theta}. \quad (13)$$

So, for $\theta \sim -\pi/2$,

$$\psi = 2 e^{ika \cos \phi}. \quad (14)$$

This is just the geometrical optics approximation. The magnetic field induced on the surface is given approximately by twice the tangential component of the incident magnetic field.

For θ near zero we have, at the stationary phase point, the condition that $\nu \sim ka$ and the Langer form reduces to the Nicholson asymptotic form⁵ which we write as⁵

$$H_r^{(1)}(ka) = \frac{-i}{\sqrt{\pi}} (ka/2)^{-1/3} w(t), \quad (15)$$

where

$$t = (ka/2)^{-1/3}(\nu - ka), \quad (16)$$

and $w(t)$ is the Airy integral

$$w(t) = \frac{1}{\sqrt{\pi}} \int_{\Gamma} e^{iZ - (1/3)Z^3} dZ \quad (17)$$

with the contour Γ given in Fig. 2. Changing the variable of integration to t ,

$$I(\theta) = -\pi i m^3 e^{ika\theta} \left\{ \frac{1}{\sqrt{\pi}} \frac{e^{im\theta t}}{w'(t)} dt \right\}, \quad (18)$$

where $m = (ka/2)^{1/3}$. The integral to be evaluated is then of the form

$$g(\xi) = \frac{1}{\sqrt{\pi}} \int \frac{e^{i\xi t}}{w'(t)} dt, \quad (19)$$

where we restrict ξ to be near zero or positive. For $\xi > 0$, $g(\xi)$ can be given as a residue series

$$g(\xi) = \frac{1}{\sqrt{\pi}} (2\pi i) \sum \frac{e^{i\xi t_n}}{\left[\frac{\partial w'(t)}{\partial t} \right]_{t=t_n}}, \quad (20)$$

where

$$w'(t_n) = 0, \quad (21)$$

and we have deformed the contour to encircle the zeros of w' which lies in the first quadrant.⁵ For $\xi \sim 0$, however, the residue series converges slowly for $\xi > 0$ and diverges for $\xi < 0$; hence $g(\xi)$ need be found by quadratures.

Substituting in (7)

$$\psi_S = \sum_{n=0}^{\infty} \{ e^{ikS_n} g(\xi_n) + e^{ikS'_n} g(\xi'_n) \}, \quad (22)$$

where we have put

$$\begin{aligned} S_n &= ka(\theta + 2\pi n) \\ S'_n &= ka(\theta' + 2\pi n), \end{aligned} \quad (23)$$

and

$$\begin{aligned} \xi_n &= (ka/2)^{1/3}(\theta + 2\pi n) \\ \xi'_n &= (ka/2)^{1/3}(\theta' + 2\pi n). \end{aligned} \quad (24)$$

We anticipate the generalization of this approach and note that S_n and S'_n are path lengths on the cylinder surface while ξ_n and ξ'_n are certain reduced distances corresponding to these path lengths.

Finally, we give the interpretation of the terms of (22) as "creeping waves." We note that the angles θ and θ' measure the angular distance from the geometrical shadow boundaries of cylinder. This is illustrated in Fig. 3. The interpretation, first proposed by Franz and

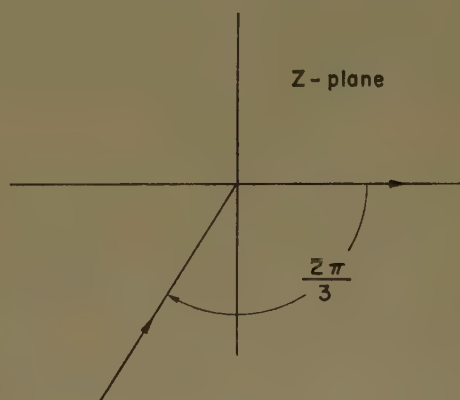


Fig. 2—The contour Γ .

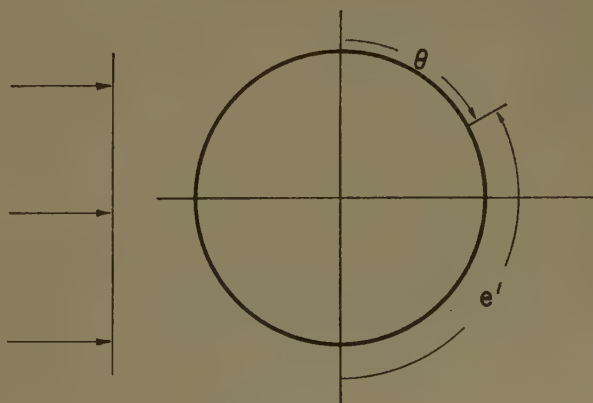


Fig. 3.

⁴ B. L. Van Der Waerden, "On the method of saddle points," *Appl. Sci. Res.*, vol. B2, pp. 33-45; 1951.

⁵ V. A. Fock, "Diffraction of radio waves around the earth's surface," *J. Phys.*, vol. 9, pp. 255-266; 1945.

Depperman,⁶ is that a wave is launched at the shadow boundary and then creeps into the shadow. The subsequent terms in the series with θ replaced by $\theta + 2\pi n$ will then represent terms which have made n circuits around the cylinder. The justification of this interpretation has been given by Friedlander.⁷

For the incident magnetic field along the y axis, we have an analogous treatment. Briefly, we have on the surface,

$$H_s = \hat{\phi} X, \quad (25)$$

where $\hat{\phi}$ is a unit vector on the surface in the ϕ direction and

$$X = \frac{1}{\pi m^3} \sum \int dv \frac{e^{iv(\theta+2\pi n)} + e^{iv(\theta'+2\pi n)}}{H_v^{(1)}(ka)}. \quad (26)$$

In and near the shadow we approximate the Hankel function by the Airy integral in (26) and find that

$$X = \frac{i}{m} \sum_n e^{ikS_n} f(\xi_n),$$

where S_n and ξ_n are as defined above and

$$f(\xi) = \frac{1}{\sqrt{\pi}} \int \frac{e^{it\xi}}{w(t)} dt. \quad (27)$$

Again, $f(\xi)$ can be evaluated by a residue series for $\xi > 0$ and by quadrature for $\xi \sim 0'$.

For $\xi < 0$, i.e., the optics region, the Nicholson approximation to the Hankel function is no longer valid. We use Langer's asymptotic approximation evaluated by stationary phase and find that

$$X \sim 2 \sin \theta e^{ikx}. \quad (28)$$

The "creeping wave" interpretation obtains just as before.

We now give Fock's work.^{8,9} Fock, by means of a physical argument, gives a description of the field in the region of the geometric shadow boundary near the surface in terms of a parabolic differential equation. The import of this in terms of Franz's concept of creeping waves will be made clear below.

We let $f(X, Y) = 0$ be the equation of a convex cylindrical surface, the cylinder axis in the Z direction. We consider a plane electromagnetic wave to be incident in the X direction and we take the origin of coordinates to be on the surface at the geometrical shadow boundary; the coordinates are given by the solution of $\partial f / \partial x = 0$.

⁶ W. Franz and K. Depperman, "Theory of diffraction by a cylinder as affected by the surface wave," *Ann. der Phys. (Lpz.)*, vol. 10, pp. 361-373; June, 1952.

⁷ F. G. Friedlander, "Diffraction of pulses by a circular cylinder," *Commun. Pure and Appl. Math.*, vol. 7, pp. 705-732; November, 1954.

⁸ V. A. Fock, "The field of a plane wave near the surface of a conducting body," *J. Phys.*, vol. 10, pp. 399-409; 1946.

⁹ J. R. Wait, "Electromagnetic Radiation from Cylindrical Structures," Pergamon Press, Inc., London, Eng., ch. 18, 1953. This chapter gives an analogous two-dimensional treatment of Fock's theory.

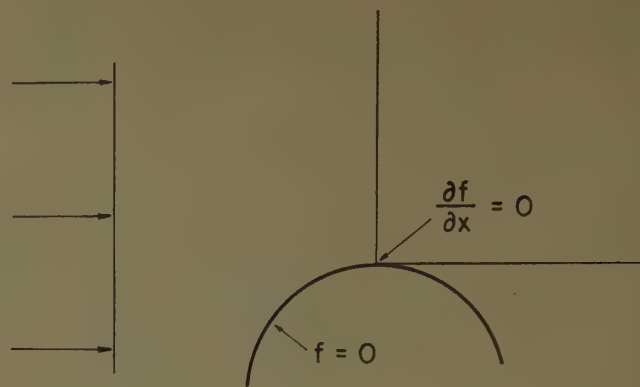


Fig. 4.

This is illustrated in Fig. 4. Further, we assume a parabolic approximation to the surface; i.e.,

$$f(x, y) = y + \frac{1}{2} \frac{x^2}{R_0}, \quad (29)$$

where R_0 is the radius of curvature at the shadow boundary, the origin of our coordinate system.

This essentially two-dimensional vector problem⁹ can be characterized in terms of the scalar problems

$$(\nabla^2 + k^2)\psi = 0, \quad (30)$$

where we assume the time dependence e^{-ikt} . The incident wave will have the functional dependence e^{ikx} , and this we introduce explicitly putting

$$\psi = e^{ikx} U. \quad (31)$$

Substituting in (30), we have

$$U_{xx} + U_{yy} + 2ikU_x = 0. \quad (32)$$

Now we come to Fock's order argument. He supposes that the variation of the functions U in the y direction, normal to the surface, is greater than the variation in the x direction. The physical content of this argument is apparent: there is a large variation in the field quantities on crossing the shadow boundary, $y=0$, $x>0$, but having eliminated the dependence on the incident field, the variation in the x direction should be relatively slow.

Fock makes the precise assumptions

$$\frac{\partial U}{\partial y} = 0 \left(\frac{k}{m} U \right), \quad (33)$$

$$\frac{\partial U}{\partial x} = 0 \left(\frac{k}{M} u \right), \quad (34)$$

where m and M are dimensionless parameters satisfying the inequalities

$$M \gg m \gg 1. \quad (35)$$

Basing our action on this order argument, we neglect the second derivative with respect to x in (30) and write

$$U_{yy} + 2ikU_x = 0. \quad (36)$$

This implies M is of order m^2 , so we put

$$M = m^2, \quad (37)$$

and define the new variables

$$\xi = \frac{mx}{R_0}, \quad (38)$$

$$\eta = \frac{2m^2}{R_0} \left(y = \frac{1}{2} \frac{x^2}{R_0} \right). \quad (39)$$

Making the change in variables, (32) becomes

$$U_{\eta\eta} + i \frac{kR_0}{2} \frac{1}{m^3} (U_\xi + \xi U_\eta) = 0. \quad (40)$$

Now, choose m such that the coefficient of the last terms becomes one; *i.e.*, put

$$m^3 = \frac{kR_0}{2}. \quad (41)$$

Eq. (40) is now

$$U_{\eta\eta} + i(U_\xi + \xi U_\eta) = 0. \quad (42)$$

For the purposes of a formal simplification we put

$$U = e^{i(\xi^2/3) - i\xi\eta} V. \quad (43)$$

This results in the equation

$$V_{\eta\eta} + \eta V + iV_\xi = 0. \quad (44)$$

If the incident magnetic field lies in the Z direction,

$$\mathbf{H}_0 = e^{ikz} \hat{\mathbf{z}}. \quad (45)$$

The total field will be of the form

$$\mathbf{H} = \psi \hat{\mathbf{z}},$$

where ψ satisfies (30) and the boundary condition

$$\left. \frac{\partial \psi}{\partial n} \right|_{f=0} = 0; \quad (46)$$

i.e., the normal derivative of ψ vanishes on the surface. In terms of the function V and the variables ξ , η , this condition is

$$\frac{\partial V}{\partial \eta} = 0 \quad \text{at} \quad \eta = 0. \quad (47)$$

A particular solution is given by

$$V = e^{i\xi\eta} w(t - \eta), \quad (48)$$

where $w(t)$ is an Airy integral, the solution of

$$w''(t) = tw(t). \quad (49)$$

Since two independent solutions are needed, we define

$$\begin{aligned} w_1(t) &= \frac{1}{\sqrt{\pi}} \int_{\Gamma_1} e^{Zt - (1/3)Z^3} dZ \\ w_2(t) &= \frac{1}{\sqrt{\pi}} \int_{\Gamma_2} e^{Zt - (1/3)Z^3} dZ, \end{aligned} \quad (50)$$

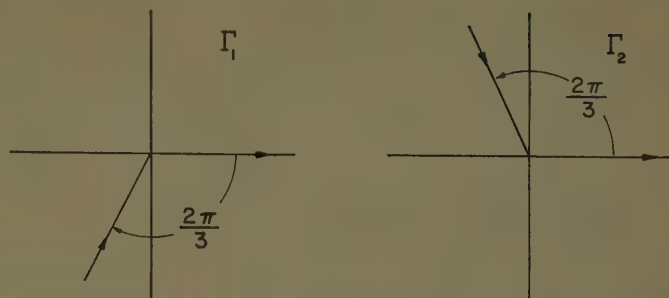


Fig. 5.

where the contours Γ_1 and Γ_2 are given in Fig. 5. We then look for the solution in the form

$$V = \frac{i}{2\sqrt{\pi}} \int_C e^{i\xi t} \left\{ w_2(t - \xi) - \frac{w_2'(t)}{w_1'(t)} w_1(t - \xi) \right\} dt, \quad (51)$$

where we note that this satisfies the differential equation and the boundary condition, and has, from the asymptotic values of the function w_1 and w_2 , the correct phase for this problem. The contour C is the same as Γ_2 in Fig. 5.

The magnetic field is then

$$HZ = e^{ikx} e^{-i\xi\zeta} + i \frac{\xi^3}{3} V, \quad (52)$$

or, making use of the relationship

$$w_1'(t)w_2(t) - w_1(t)w_2'(t) = -2i, \quad (53)$$

the field on the surface, $\zeta=0$, is given by

$$H_Z = e^{ikx} G(\xi), \quad (54)$$

where

$$G(\xi) = e^{i\xi^3/3} \frac{1}{\sqrt{\pi}} \int_C dt \frac{e^{i\xi t}}{w'(t)}. \quad (55)$$

For the incident field

$$\mathbf{H}_0 = ye^{ikx}, \quad (56)$$

we have the boundary conditions

$$H_Z = 0, \quad (57)$$

and, since we confine our attention to the region of the shadow boundary, the normal component of the magnetic field is given by H_y .

$$H_y = 0 \quad \text{on} \quad S. \quad (58)$$

Writing as before,

$$H_y = e^{ikx} \Phi$$

must satisfy

$$\nabla^2 \Phi + 2ik \frac{\partial \Phi}{\partial x} = 0, \quad (59)$$

and

$$\Phi = 0 \quad \text{on} \quad S.$$

We repeat the order argument and write (59) as

$$\frac{\partial^2 \Phi}{\partial y^2} + 2ik \frac{\partial \Phi}{\partial x} = 0 \quad (60)$$

$$\Phi = 0 \quad \text{on } S. \quad (61)$$

Now \mathbf{H} is a divergence-free field:

$$\nabla \cdot \mathbf{H} = 0, \quad (62)$$

and the divergence condition becomes

$$ikH_x^* + \nabla \cdot \mathbf{H}^* = 0. \quad (63)$$

Repeating the order argument we neglect $\partial H_x^* / \partial x$, as compared with ikH_x^* , leaving

$$H_x^* = \frac{i}{k} \frac{\partial H_y^*}{\partial y}, \quad (64)$$

or

$$H_x^* = \frac{i}{m} \frac{\partial \Phi}{\partial \eta}. \quad (65)$$

The field component H_x on the surface is given by

$$H_x = e^{ikx} \frac{i}{m} e^{-i\xi\eta + i(\xi^3/3)} \frac{\partial \psi}{\partial \eta}, \quad (66)$$

where ψ satisfies

$$\begin{aligned} \psi_{\eta\eta} + \eta\psi + i\psi_\xi &= 0 \\ \psi &= 0 \quad \text{on } S. \end{aligned} \quad (67)$$

This has a particular solution:

$$e^{i\xi t} w(t - \eta). \quad (68)$$

So, we write

$$\psi = \frac{1}{2\sqrt{\pi}} \int e^{i\xi t} \left\{ w_2(t - \eta) - \frac{w_2(t)}{w_1(t)} w_1(t - \eta) \right\} dt, \quad (69)$$

and

$$\left. \frac{\partial \psi}{\partial \eta} \right|_{\eta=0} = \frac{1}{\sqrt{\pi}} \int \frac{e^{i\xi t}}{w(t)} dt. \quad (70)$$

Using the notation

$$f(\xi) = \frac{1}{\sqrt{\pi}} \int \frac{e^{i\xi t}}{w(t)} dt, \quad (71)$$

we have, on the surface,

$$H_x = \frac{i}{m} e^{i(\xi^3/3)} e^{ikx} f(\xi). \quad (72)$$

We are now able to apply Fock's solution to the circular cylinder and compare it with the Franz "creeping wave" solution. From (20) and (55) we have, for the circular cylinder solution and the Fock solution, the function

$$g(\xi) = \frac{1}{\sqrt{\pi}} \int \frac{e^{i\xi t}}{w'(t)} dt. \quad (73)$$

However, we note that for the circular cylinder

$$\xi = \left(\frac{ka}{2} \right)^{1/3} \theta, \quad (74)$$

while in the Fock treatment

$$\xi_F = \left(\frac{kR_0}{2} \right)^{1/3} \frac{x}{R_0}, \quad (75)$$

or, in terms of polar coordinates, since $R_0 = a$ and $x = a \sin \theta$,

$$\xi_F = \left(\frac{ka}{2} \right)^{1/3} \sin \theta. \quad (76)$$

These arguments of the functions $g(\xi)$ thus agree to first order for $\theta \sim 0$. This imposes a restriction on the applicability of Fock's method as it stands.

To bring these solutions into agreement we return to the creeping wave interpretation and Fock's derivation of the parabolic differential equation. From the creeping wave interpretation we have a wave launched at the boundary which then creeps *along* the surface into the shadow boundary. Now the natural description of such a phenomenon would be by a parabolic differential equation. This Fock has done. However, we note that ξ_F measures distance along the direction of propagation rather than along the surface of the obstacle.

We now observe that the argument used by Fock in his derivation of the parabolic equation is also applicable outside the region $\xi \sim 0$, provided we compare the variations of field along the surface of the obstacle and perpendicular to the surface of the obstacle; *i.e.*, we can use the Fock equation anywhere in the shadow region provided we define a new set of variables ξ and η for each increment we move into the shadow. To illustrate this we write the formal solution of (44) as

$$V(\xi, \eta) = e^{-iT\xi} V(0, \eta), \quad (77)$$

where

$$T = - \left(\frac{\partial^2}{\partial \eta^2} + \eta \right).$$

This gives us an expression valid for $\xi < \xi_1 \gg 1$, for example. Given this, we then redefine our variables and write

$$V(\xi, \eta) = e^{-iT(\xi - \xi_1)} V(\xi_1, \zeta), \quad (78)$$

which generalizes to

$$V(\xi, \eta) = e^{-iT\xi} V(\xi, \zeta), \quad (79)$$

with

$$\xi = \int_0^S \left(\frac{kR(s)}{2} \right)^{1/3} \frac{ds}{R(s)}, \quad (80)$$

where ds is the element of path length along the surface and $R(s)$ is the radius of curvature at s .

Applying this reasoning to the circular cylinder, we have

$$\xi = \int_0^\theta \left(\frac{ka}{2}\right)^{1/3} \frac{a d\theta}{a} \quad (81)$$

$$= \left(\frac{ka}{2}\right)^{1/3} \theta. \quad (82)$$

This is, however, the expression appearing in Franz's treatment and, hence, we have brought the Franz and Fock solutions into agreement.

This important generalization of Fock's work was systematically used by J. B. Keller.¹⁰ Keller proceeds from a local solution of the circular cylinder similar to the above treatment.

THREE-DIMENSIONAL PROBLEMS

We now turn to the application of Fock's method to three-dimensional problems. There is an essential complication present in the case of finite, convex, three-dimensional surfaces. Again we will illustrate the general problem by a prototype problem, *i.e.*, the scalar scattering by a sphere.¹¹

We start with the Dirichlet boundary condition; *i.e.*, we wish the solution of

$$(\nabla^2 + k^2)\psi = 0 \quad (83)$$

$$\psi(a) = 0 \quad (84)$$

where a is the radius of the sphere. In particular, we want to determine

$$\left. \frac{\partial \psi}{\partial r} \right|_{r=a} \quad (85)$$

Let the incident field approach along the polar axis. Then the normal derivative of the field induced on the surface of a sphere of radius a is given by the series

$$\frac{\partial \psi_0}{\partial n} = \sum \left(n + \frac{1}{2}\right) P_n(\cos \theta) e^{-i(\pi/2)n} \frac{1}{\zeta_n^{(1)}(ka)}. \quad (86)$$

Since the summand has no singularities with respect to the index on the positive real axis, this can be written as the contour integral

$$\frac{\partial \psi_0}{\partial n} = \dots \int_C \nu e^{-i(\pi/2)\nu} \frac{1}{\zeta_{\nu-1/2}^{(1)}(ka)} \sec \nu \pi P_{\nu-1/2}^*(\cos \theta) d\nu \quad (87)$$

where C is a contour encircling the positive real axis, as in Fig. 6, and where we use the notation

$$P_\nu^*(x) = P_\nu(-x). \quad (88)$$

¹⁰ J. B. Keller, "Diffraction by a convex cylinder," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-4, pp. 312-321; July, 1956.

¹¹ Corresponding vector problems are treated in V. A. Fock, M. G. Belkina, and L. A. Weinstein, "Diffraction of Electromagnetic Waves on Certain Bodies of Revolution," Soviet Radio Press, Moscow, USSR; 1958.

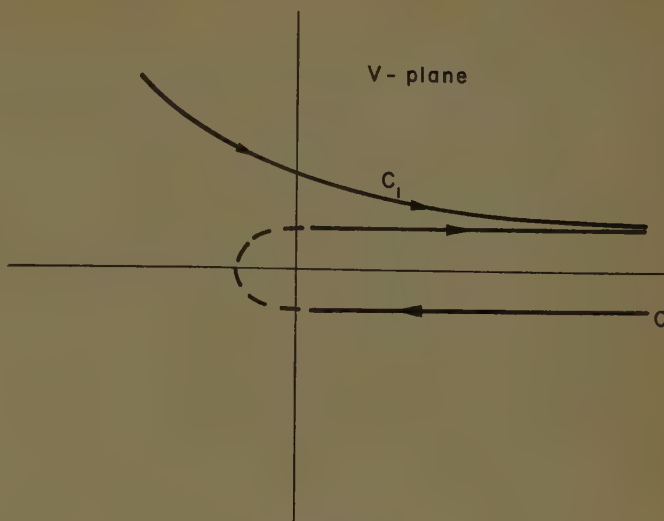


Fig. 6.

Since the integrand is an odd function of ν regular in the second and fourth quadrants, and having simple poles [the zeros of $\zeta_{\nu-1/2}^{(1)}(ka)$ in the first quadrant] we change the contour C to C_1 . Since $\text{Im} \nu > 0$ along C_1 , we make the convergent expansion

$$\sec \nu \pi = e^{i\pi\nu} \sum_{n=0}^{\infty} e^{2\pi i\nu n} (-)^n. \quad (89)$$

Now, making use of the reduction of the Legendre function, we have

$$\sec \nu \pi P_{\nu-1/2}^*(\cos \theta) = e^{i(\pi/2)\nu} \sum (-)^n \cdot \{P_{\nu-1/2}^{(+)}(\theta + 2\pi n) - P_{\nu-1/2}^{(+)}(2\pi(n+1) - \theta)\}. \quad (90)$$

Substituting in the integrand,

$$I(\theta) = e^{i(\pi/2)\nu} \sum (-)^n \cdot \{I_1(\theta + 2\pi n) - I_1(2\pi(n+1) - \theta)\} \quad (91)$$

where

$$I_1(\theta) = \int_{C_1} d\nu \frac{e^{-i(\pi/2)\nu}}{\zeta_{\nu-1/2}^{(1)}(ka)} P_{\nu-1/2}^{(+)}(\theta). \quad (92)$$

For the range $\pi/6 < \theta < 5\pi/6$, we can use the expansion

$$P_{\nu-1/2}^{(+)}(\theta) = \frac{1}{\sqrt{2i \sin \theta}} \frac{\Gamma(\nu + \frac{1}{2})}{\Gamma(\nu + 1)} \cdot e^{i\nu\theta} F\left(\frac{1}{2}, \frac{1}{2}; \nu + 1; \frac{e^{i\theta}}{2i \sin \theta}\right) \quad (93)$$

and the notation

$$\phi = \theta - \frac{\pi}{2}, \quad \phi' = \frac{3\pi}{2} - \theta \quad (94)$$

so that

$$I(\theta) = \frac{e^{i(\pi/2)}}{\sqrt{2\pi i \sin \theta}} \sum (-)^n \left\{ \int_{C_1} \frac{\nu \Gamma(\nu + \frac{1}{2})}{\Gamma(\nu + 1)} \frac{1}{\zeta_{\nu-1/2}^{(1)}(ka)} \right. \\ \cdot \left(e^{i\nu(\phi+2\pi n)} F\left(\frac{1}{2}, \frac{1}{2}; \nu + 1; \frac{e^{i\phi}}{2 \cos \phi}\right) \right. \\ \left. \left. - i e^{i\nu(\phi'+2\pi n)} F\left(\frac{1}{2}, \frac{1}{2}; \nu + 1; \frac{e^{i\phi'}}{2 \cos \phi'}\right) \right) \right\}. \quad (95)$$

Or defining

$$I_2(\phi) = \int_{C_1} d\nu \frac{\Gamma(\nu + \frac{1}{2})}{\Gamma(\nu + 1)} \frac{1}{\zeta_{\nu-1/2}^{(1)}(ka)} \\ \cdot e^{i\nu\phi} F\left(\frac{1}{2}, \frac{1}{2}; \nu + 1; \frac{e^{i\phi}}{2 \cos \phi}\right) \quad (96)$$

we have

$$I(\phi) = \frac{e^{i(\pi/2)}}{\sqrt{2\pi i \sin \theta}} \sum_{n=0}^{\infty} (-)^n I_2(\phi + 2\pi n) \\ - i I_2(\phi' + 2\pi n). \quad (97)$$

The terms of this series correspond to the "creeping waves" of Franz. In this sense we note that ϕ and ϕ' are just the angular distances measured from the shadow boundaries. (See Fig. 7.) The field induced on the surface

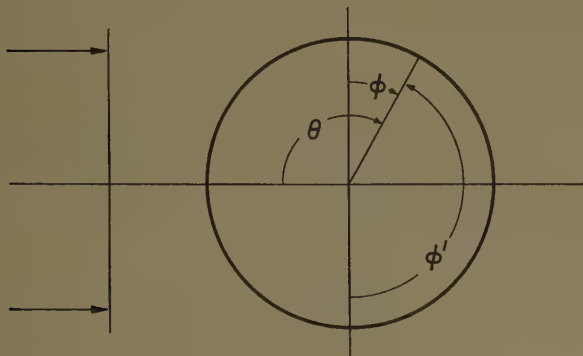


Fig. 7.

under the imposition of the Neumann boundary condition follows immediately from the above on the substitution of the derivative of the spherical Hankel function in the denominator of the summand or integrand. We then treat with integrals of the form

$$I_2'(\phi) = \int_{C_1} d\nu \frac{\nu \Gamma(\nu + \frac{1}{2})}{\Gamma(\nu + 1)} \frac{1}{\zeta_{\nu-1/2}^{(1)}(ka)} e^{i\nu\phi} F. \quad (98)$$

We consider the integrals (96) and (98) in more detail. Assume a sufficiently large value of ka so that the Airy integral approximation for the Hankel function and its derivative is justified. We put, for example,

$$\zeta_{\nu-1/2}^{(1)}(ka) \cong -im^{1/2}w(t) \quad (99)$$

where

$$m = \left(\frac{ka}{2}\right)^{1/3}, \quad t = \frac{\nu - ka}{m}. \quad (100)$$

The integral I_2 is then, essentially, of the form

$$I_2 = \int_{C_2} \phi(t) \frac{e^{i\xi t}}{w(t)} dt \quad (101)$$

where we put $\xi = m\phi$, and C_2 is the contour running from infinity along $\arg t = 2\pi/3$ to the origin and from the origin to infinity along $\arg t = 0$.

We note that the form of (101) is very like that of Fock's function. In fact, Fock, in his paper,⁵ arrives at just this form, which he then approximates by using the asymptotic form

$$\nu \frac{\Gamma(\nu + \frac{1}{2})}{\Gamma(\nu + 1)} F\left(\frac{1}{2}, \frac{1}{2}; \nu + 1; \frac{e^{i\theta}}{2i \sin \theta}\right) \sim \sqrt{\nu}. \quad (102)$$

Then, in the shadow region, he evaluates (101) as a residue series. Since the first pole occurs near $\nu = ka$, and this for $|ka \sin \theta| \gg 1$ is the principal contributor to the residue series, he approximates (102) by \sqrt{ka} . That which remains is precisely one of Fock's functions.

Returning to (95), we see that there is an essential difference between the sphere results and the parabolic equation results. The physical significance of this difference is seen immediately on noting that the parabolic equation is strictly applicable to a two-dimensional problem (the infinite circular cylinder) while the sphere, being a finite body, forces the waves creeping into the shadow to converge on the pole $\theta = \pi$. This accounts for the term $1/\sqrt{\sin \theta}$ in (95). In fact, if we consider that the energy surface density must increase inversely as the available space, we have

$$E \sim |\psi|^2 \sim \frac{1}{\sin \theta}, \quad (103)$$

or

$$\psi \sim \frac{1}{\sqrt{\sin \theta}}. \quad (104)$$

This result has already been noted by Franz.¹

N. A. Logan¹² has applied Fock's reasoning in approximating the asymptotic form of (101) itself. Since the major contribution to the integral comes from the region $t \sim 0$, and $\phi(t)$ is a slowly varying function, he writes

$$I_2 = \left\{ \int_{C_2} \frac{e^{i\xi t}}{w(t)} dt \right\} \phi(0) \\ = \frac{ka \Gamma(ka + \frac{1}{2})}{\Gamma(ka + 1)} F\left(\frac{1}{2}, \frac{1}{2}; ka + 1; \frac{e^{i\theta}}{2i \sin \theta}\right) \\ \cdot \int_{C_2} \frac{e^{i\xi t}}{w(t)} dt, \quad (105)$$

¹² N. A. Logan, private communication.

where he restricts the region of applicability to

$$|ka \sin \theta| \gg 1.$$

Near the pole $\theta = \pi$ we use the asymptotic representation,

$$P_\nu^*(\theta) = J_0 \left((2\nu + 1) \cos \frac{\theta}{2} \right) + O \left(\cos^2 \frac{\theta}{2} \right), \quad (106)$$

rather than the decomposition into $P_\nu^{(+)}$ and $P_\nu^{(-)}$ which are singular at $\theta = \pi$.

Using the above approximations, Mr. Logan has made a comparison with the sum of the harmonic series for $ka = 20$ and has found that the agreement between the two is good. In the transition region between $|\nu \sin \theta| \gg 1$ and $|\sin \theta| \gg 1$ the results from either side are continued into this region, and even here the agreement is also good.

Fock⁸ has applied his method to the three-dimensional problem. His result is precisely the same as that presented above for the two-dimensional problem.

We consider a finite, smooth, convex, perfectly-conducting body illuminated by a plane electromagnetic wave. We take the plane wave to be incident along the x direction and erect a coordinate system at some point on the shadow boundary with the y axis normal to the surface and the z axis chosen so as to form a right-handed system, as shown in Fig. 8.

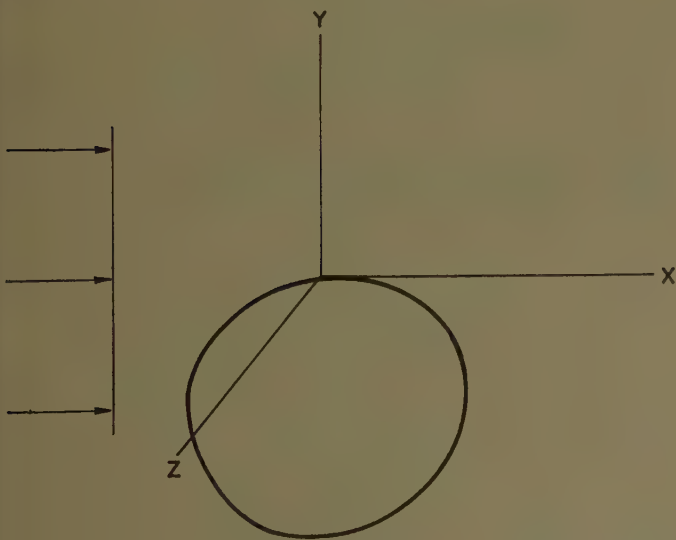


Fig. 8.

As before, we perform a local analysis near the origin of the coordinate system. We write

$$\begin{aligned} H &= e^{ikx} H^* \\ E &= e^{ikx} E^*. \end{aligned} \quad (107)$$

Hence, the starred quantities satisfy

$$\begin{aligned} \nabla_x H^* + ik\hat{x} \times H^* &= -ikE^* \\ \nabla_x E^* + ik\hat{x} \times E^* &= ikH^*, \end{aligned} \quad (108)$$

from Maxwell's equations for free space.

Now we extend the order argument and write, letting ψ stand for any of the field components in (108),

$$\begin{aligned} \frac{\partial \psi}{\partial y} &= O \left(\frac{k}{m} \psi \right), \\ \frac{\partial \psi}{\partial x} &= O \left(\frac{k}{M} \psi \right), \\ \frac{\partial \psi}{\partial z} &= O \left(\frac{k}{M} \psi \right), \end{aligned} \quad (109)$$

where M and m satisfy (35) and (37) above.

Since each of the starred field components must satisfy the reduced wave equation

$$\nabla^2 \psi + 2ik \frac{\partial \psi}{\partial x} = 0, \quad (110)$$

by applying the ordering assumption of (110) we have

$$\frac{\partial^2 \psi}{\partial y^2} + 2ik \frac{\partial \psi}{\partial x} = 0. \quad (111)$$

Finally, dropping the asterisks, we have from the ordering assumptions and (108)

$$\begin{aligned} E_x &= \frac{i}{k} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}, \\ E_y &= H_x, \\ E_z &= -H_y, \\ H_x &= \frac{i}{k} \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z}. \end{aligned} \quad (112)$$

We see, then, that the three-dimensional solutions of Fock are precisely the two-dimensional solutions given above. In fact, for the incident field

$$H_0 = e^{ikx\hat{x}}, \quad (113)$$

we put

$$H_x = e^{ikx}\psi. \quad (114)$$

Then, near the shadow boundary, ψ must satisfy

$$\frac{\partial^2 \psi}{\partial y^2} + 2ik \frac{\partial \psi}{\partial x} = 0, \quad (115)$$

and the boundary condition

$$\left. \frac{\partial \psi}{\partial n} \right|_S = 0. \quad (116)$$

These are precisely the conditions on H_x in the two-dimensional problem as given above in (46).

For the incident field

$$H_0 = e^{ikx}\hat{y}. \quad (117)$$

On the other hand, we put

$$H_y = e^{ikx}\phi, \quad (118)$$

where ϕ satisfies

$$\frac{\partial^2 \phi}{\partial y^2} + 2ik \frac{\partial \phi}{\partial x} = 0, \quad (119)$$

$$\left. \phi \right|_s = 0. \quad (120)$$

These are the conditions given in (59) for the two-dimensional problem. Again making use of the divergence condition, we have

$$H_x = e^{ikx} \frac{i}{k} \frac{\partial \phi}{\partial y}. \quad (121)$$

We emphasize the fact that the application of this essentially two-dimensional approach is restricted to the region of the geometric shadow boundary. To carry these solutions farther into the shadow region, we must make use of the fact that as the surface area decreases as it goes into the shadow region, the energy density must increase. Further, we make the point that the "creeping waves" propagate along geodesics on going into the shadow. The first requirement was illustrated in the treatment of the scalar sphere problem in the appearance of the factor $(\sin \theta)^{-1/2}$ in the expression for the field. The second requirement was met in the tacit assumption that the creeping waves followed great circles on the sphere.

To determine more generally the convergence factor corresponding to $(\sin \theta)^{-1/2}$ in the case of the sphere, we consider two adjoining geodesic paths arising on the geometrical shadow boundary. The geodesics are determined by two conditions:

- 1) the point on the shadow boundary at which they arise; and

- 2) the angle which the incident radiation makes with the shadow curve.

We write for the two paths

$$\begin{aligned} r_1 &= r(l, S_1), \\ r_2 &= r(l + \Delta l, S_2), \end{aligned} \quad (122)$$

where l and $l + \Delta l$ are points on the shadow boundary and S_1 and S_2 are path lengths along the geodesics. This is shown in Fig. 9. We choose S_1 and S_2 so that r_1 and

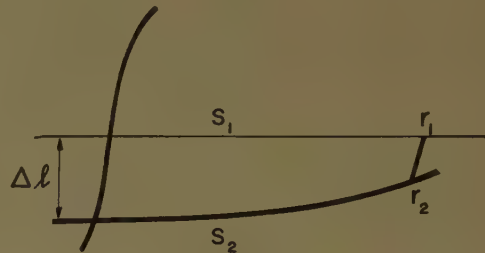


Fig. 9.

r_2 be equiphase points; then the convergence of area available to the energy propagating into the shadow will be proportional to

$$A = \left| \frac{r_1 - r_2}{\Delta l} \right|, \quad (123)$$

or, in the limit as Δl vanishes,

$$A(l, S) = \left| \frac{dr(S, l)}{dl} \right|. \quad (124)$$

In order to use the Fock method, then, we require the field functions to be multiplied by the factor $A^{-1/2}$.

Reduction of the Integral Equations for High-Frequency Diffraction by Disks and Strips*

B. NOBLE†

Summary—The kernels of the integral equations for scalar diffraction by strips and disks are special cases of a kernel connected with the generalized axially symmetrical wave equation. A transformation of this kernel enables the original singular integral equations to be reduced to Fredholm integral equations of the second kind. These can be solved asymptotically at high frequencies.

Applications are made to diffraction by strips and disks with incident waves of arbitrary form. Special results involving diffraction of plane waves are recovered from the general formulas.

I. INTRODUCTION

SO MUCH has been written on high-frequency diffraction by strips and disks in recent years that the appearance of still one more paper on the subject would seem to require some justification.

It is well known that at high frequencies the most convenient solution is an asymptotic expansion in terms of inverse powers of ka , where k is the wave number and a is a characteristic dimension (the radius of the disk or the half-width of the strip). The classical theory of Kirchhoff gives a first approximation but it is difficult to extend the theory to obtain higher orders of approximation. A review is given by Bouwkamp [2]. The simplest way of obtaining the asymptotic expansion is undoubtedly Keller's method of geometrical acoustics which has been applied to strip and disk problems [10]. It is of considerable interest, however, to obtain the expansion by other means.

Since diffraction problems involving disks and strips can be formulated simply and elegantly in terms of integral equations it would seem natural to try to obtain directly the asymptotic solution from the integral equations. This is the aim of the present paper.

In Section II we consider a general kernel which includes the kernels for strip and disk as special cases. Transformation of this kernel allows us to convert the original singular integral equations into Fredholm integral equations of the second kind from which the asymptotic expansions can be obtained.

Most of the literature relevant to the present paper deals with two-dimensional problems involving strips or slits. The first rigorous treatment of diffraction by a slit at high frequencies was given by Schwartzchild [18] who suggested a method of successive approximations, starting from the Sommerfeld half-plane solution. A related integral equation approach was given by Fox [7] who reduced the integral equation for pulse diffraction by a slit to a system of Fredholm integral equations of

the second kind. The analogous steady-state equations (Baker and Copson [1]) have been used by Millar [15] to obtain in some detail an asymptotic expansion for high frequency diffraction by a slit. The Fredholm integral equations used in this work have been rediscovered recently by Grinberg [8]. Similar integral equations have been obtained for a two-dimensional wing problem in aerodynamics by Eckhaus [5]. The present paper has points of contact with all these references.

A closely related approach to diffraction by strips is to use an approximate method based on the Wiener-Hopf technique. Again the essential feature is a method of successive approximations, starting from the half-plane solution. This has been examined by Levine [12], using an integral equation method, and by Noble,¹ using a method due to Jones.

Comparatively little has been done to extend these ideas to diffraction by disks. Levine [11] and Levine and Wu [13] have given an interesting approximate solution of the integral equations for disk diffraction, using the Wiener-Hopf technique. One aim of the present work is to develop a corresponding extension of the equations of Fox, *et al.*, for disk diffraction problems.

There are various other relevant references. For slit diffraction there is the perturbation or interaction method of Karp [9], [20]. An "edge-current" idea has been applied to slits by Clemmow [4] and to disks by Millar [14]. A hyperbolic equation method has been applied to slit diffraction by Burger [3] and Timman [19]. A generalization of the decomposition into repeated Volterra integral equations in Section III of this paper is given by Noble,² as is a generalization of the Fredholm integral equation of Fox, *et al.*³

II. TRANSFORMATION OF A KERNEL

The transformation obtained in this section is basic for developments in the remainder of the paper. The kernel which we consider is

$$K(\rho, r) = \int_0^\infty \frac{t}{\gamma} J_\nu(\rho t) J_\nu(r t) dt, \quad (1)$$

where

$$\gamma = (t^2 - k^2)^{1/2} \text{ for } t \geq k, \quad -i(k^2 - t^2)^{1/2} \text{ for } t \leq k. \quad (2)$$

The change of variable $t = k\xi$ in (1) gives

$$K(\rho, r) = k \int_0^\infty \frac{\xi}{(\xi^2 - 1)^{1/2}} J_\nu(k\rho\xi) J_\nu(kr\xi) d\xi. \quad (3)$$

* The research reported in this paper has been sponsored by the AF Cambridge Research Center, Air Res. and Dev. Command, under Contract No. AF 19(604)5238.

† New York University, New York, N. Y. On leave of absence from the Royal College of Science and Technology, Glasgow, Scotland.

¹ See for instance [16], pp. 196–207.

² *Ibid.*, p. 230, ex. 6.4.

³ *Ibid.*, p. 231, ex. 6.5.

We wish to introduce the asymptotic expansions of the Bessel functions for large k but this would be of doubtful validity at this stage since the integrals of higher order terms would be divergent near $\xi=0$. It is natural to split the infinite integral for $K(\rho, r)$ in (3) into the ranges $(0, 1)$ and $(1, \infty)$. The divergence comes from the range $(0, 1)$. This part can be transformed in the following way.

We shall work in the complex plane $\zeta = \xi + i\eta$. Suppose that $\rho > r$. Since $J_\nu(k\rho\zeta)$ tends to infinity as ζ tends to infinity in both upper and lower half-planes we write

$$J_\nu(k\rho\zeta)J_\nu(kr\zeta) = \frac{1}{2}\{H_\nu^{(1)}(k\rho\zeta)J_\nu(kr\zeta) + H_\nu^{(2)}(k\rho\zeta)J_\nu(kr\zeta)\},$$

where the first and second terms on the right tend to zero in upper and lower half-planes, respectively. Suppose that δ, ϵ are small real constants, and R is a large real constant. Integrate

$$\zeta(\zeta^2 - 1)^{-1/2}H_\nu^{(1)}(k\rho\zeta)J_\nu(kr\zeta), \quad (\rho > r),$$

round a contour consisting of $\zeta = \delta$ to $1 - \epsilon$, the semi-circle of radius ϵ above $\zeta = 1$, $\zeta = 1 + \epsilon$ to R , the quarter-circle $R \exp i\Theta$, $0 \leq \Theta \leq \pi/2$, the imaginary axis iR to $i\delta$, and the quarter-circle $\delta \exp i\theta$, $\pi/2 \geq \theta \geq 0$. There are no singularities inside this contour so that the integral is zero. If we let δ, ϵ tend to zero and R tend to infinity the contributions from the corresponding arcs tend to zero. Hence

$$\begin{aligned} & -i \int_0^1 \frac{\xi}{(1 - \xi^2)^{1/2}} H_\nu^{(1)}(k\rho\xi)J_\nu(kr\xi)d\xi \\ & + \int_1^\infty \frac{\xi}{(\xi^2 - 1)^{1/2}} H_\nu^{(1)}(k\rho\xi)J_\nu(kr\xi)d\xi \\ & + i \int_\infty^0 \frac{\eta}{(1 + \eta^2)^{1/2}} H_\nu^{(1)}(ik\rho\eta)J_\nu(ikr\eta)d\eta = 0. \quad (4) \end{aligned}$$

Integration over a similar contour in the lower right-hand quadrant, passing under the branch-point at $\zeta = 1$, gives

$$\begin{aligned} & i \int_0^1 \frac{\xi}{(1 - \xi^2)^{1/2}} H_\nu^{(1)}(k\rho\xi)J_\nu(kr\xi)d\xi \\ & + \int_1^\infty \frac{\xi}{(\xi^2 - 1)^{1/2}} H_\nu^{(1)}(k\rho\xi)J_\nu(kr\xi)d\xi \\ & - i \int_\infty^0 \frac{\eta}{(1 + \eta^2)^{1/2}} H_\nu^{(1)}(-ik\rho\eta)J_\nu(-ikr\eta)d\eta = 0. \quad (5) \end{aligned}$$

But

$$\begin{aligned} H_\nu^{(1)}(ik\rho\eta)J_\nu(ikr\eta) &= -i(2/\pi)K_\nu(k\rho\eta)I_\nu(kr\eta) \\ &= -H_\nu^{(2)}(-ik\rho\eta)J_\nu(-ikr\eta). \end{aligned}$$

Hence, on subtracting (5) from (4), for $\rho > r$,

$$\begin{aligned} & \int_0^1 \frac{\xi}{(1 - \xi^2)^{1/2}} J_\nu(k\rho\xi)J_\nu(kr\xi)d\xi \\ &= \int_1^\infty \frac{\xi}{(\xi^2 - 1)^{1/2}} Y_\nu(k\rho\xi)J_\nu(kr\xi)d\xi. \end{aligned}$$

On substituting this result in (3), since

$$(\xi^2 - 1)^{1/2} = -i(1 - \xi^2)^{1/2}$$

in (3), we have

$$K(\rho, r) = k \int_1^\infty \frac{\xi}{(\xi^2 - 1)^{1/2}} H_\nu^{(1)}(k\rho\xi)J_\nu(kr\xi)d\xi, \quad (\rho > r). \quad (6)$$

We can now substitute the asymptotic expansions for the Bessel functions. This gives the basic result

$$\begin{aligned} K(\rho, r) &= \frac{1}{\pi(\rho r)^{1/2}} \{K_0(\rho, r) + e^{-i(r+1/2)\pi}K_1(\rho, r) \\ &\quad + \text{higher order terms}\}, \quad (7a) \end{aligned}$$

where

$$K_0(\rho, r) = \int_1^\infty \frac{1}{(\xi^2 - 1)^{1/2}} e^{ik\xi|\rho-r|}d\xi, \quad (7b)$$

$$K_1(\rho, r) = \int_1^\infty \frac{1}{(\xi^2 - 1)^{1/2}} e^{ik\xi(\rho+r)}d\xi. \quad (7c)$$

We make the following change of variable in (7b):

$$\xi|\rho - r| = 2v - \rho - r.$$

Then

$$K_0(\rho, r) = e^{-ik(\rho+r)} \int_{\max(\rho, r)}^\infty \frac{e^{2ikv}}{(v - \rho)^{1/2}(v - r)^{1/2}} dv. \quad (8)$$

Similarly the change of variable

$$\xi(\rho + r) = 2v - \rho + r$$

in (7c) gives

$$K_1(\rho, r) = e^{-ik(\rho-r)} \int_\rho^\infty \frac{e^{2ikv}}{(v - \rho)^{1/2}(v + r)^{1/2}} dv. \quad (9)$$

III. REDUCTION OF AN INTEGRAL EQUATION

We shall obtain an alternative form for the integral equation

$$\begin{aligned} & \int_a^\infty f(r) \{K_0(\rho, r) + \lambda K_1(\rho, r)\} dr \\ &= - \int_0^a f(r) \{K_0(\rho, r) + \lambda K_1(\rho, r)\} dr, \quad (\rho > a), \quad (10) \end{aligned}$$

where $f(r)$ is known in $0 \leq \rho \leq a$, unknown in $\rho > a$, λ is a constant, and the kernels K_0, K_1 have been defined in (7b) and (7c). For convenience we shall combine the right- and left-hand sides of (10). Insertion of the forms (8) and (9) for K_0 and K_1 gives

$$\begin{aligned} & \int_0^\infty e^{-ikrf(r)} \int_{\max(\rho, r)}^\infty \frac{e^{2ikv}}{(v - \rho)^{1/2}(v - r)^{1/2}} dv dr \\ &+ \lambda \int_0^\infty e^{ikrf(r)} \int_\rho^\infty \frac{e^{2ikv}}{(v - \rho)^{1/2}(v + r)^{1/2}} dv dr = 0, \quad (\rho > a). \quad (11) \end{aligned}$$

If orders of integration are interchanged in the first double integral, we have

$$\begin{aligned} \int_0^\infty dr \int_{\max(\rho, r)}^\infty dv &= \int_0^\rho dr \int_\rho^\infty dv + \int_\rho^\infty dr \int_r^\infty dv \\ &= \int_\rho^\infty dv \int_\rho^\rho dr + \int_\rho^\infty dv \int_\rho^v dr = \int_\rho^\infty dv \int_0^v dr. \end{aligned}$$

Hence, (11) gives

$$\begin{aligned} \int_\rho^\infty \frac{e^{2ikv}}{(v-\rho)^{1/2}} \left\{ \int_0^v \frac{e^{-ikrf(r)}}{(v-r)^{1/2}} dr \right. \\ \left. + \lambda \int_0^\infty \frac{e^{ikrf(r)}}{(v+r)^{1/2}} dr \right\} dv = 0, \quad (\rho > a). \quad (12) \end{aligned}$$

But the solution of the Abel integral equation

$$\int_\rho^\infty \frac{J(v)}{(v-\rho)^{1/2}} dv = 0, \quad \rho > a,$$

is simply

$$J(v) = 0, \quad (\rho > a).$$

Hence from (12)

$$\begin{aligned} \int_a^v \frac{e^{-ikrf(r)}}{(v-r)^{1/2}} dr + \lambda \int_a^\infty \frac{e^{ikrf(r)}}{(v+r)^{1/2}} dr \\ = - \int_0^a \left\{ \frac{e^{-ikrf(r)}}{(v-r)^{1/2}} + \lambda \frac{e^{ikrf(r)}}{(v+r)^{1/2}} \right\} dr, \quad (v > a), \quad (13) \end{aligned}$$

where we have placed unknown terms on the left and known terms on the right. We next use the inversion formula for Abel's integral equation:

$$\begin{aligned} \int_a^v \frac{P(r)}{(v-r)^{1/2}} dr = Q(v), \quad (v > a), \\ P(r) = \frac{1}{\pi} \frac{d}{dr} \int_a^r \frac{Q(v)}{(r-v)^{1/2}} dv, \quad (r > a). \end{aligned}$$

If this is applied to (13) to express $f(r)$ from the first integral in terms of the remaining three integrals, we obtain the following Fredholm equation of the second kind for $f(r)$:

$$\begin{aligned} e^{-ik\rho f(\rho)} + \frac{\lambda}{\pi} \int_a^\infty e^{ikrf(r)} k_1(\rho, r) dr \\ = - \frac{1}{\pi} \int_0^a \{ e^{-ikrf(r)} k_0(\rho, r) + \lambda e^{ikrf(r)} k_1(\rho, r) \} dr, \\ (\rho > a), \end{aligned}$$

where, after evaluation of integrals,

$$\begin{aligned} k_0(\rho, r) &= \frac{(a-r)^{1/2}}{(\rho-a)^{1/2}(\rho-r)}, \\ k_1(\rho, r) &= \frac{(a+r)^{1/2}}{(\rho-a)^{1/2}(\rho+r)}. \end{aligned}$$

We set

$$F(\rho) = (\rho-a)^{1/2} e^{-ik\rho f(\rho)}, \quad (\rho > a).$$

Then

$$F(\rho) + \frac{\lambda}{\pi} \int_a^\infty e^{2ikr} F(r) \frac{(a+r)^{1/2}}{(r-a)^{1/2}(\rho+r)} dr = H(\rho), \quad (\rho > a),$$

where

$$\begin{aligned} H(\rho) &= - \frac{1}{\pi} \int_0^a \left\{ e^{-ikrf(r)} \frac{(a-r)^{1/2}}{(\rho-r)} \right. \\ &\quad \left. + \lambda e^{ikrf(r)} \frac{(a+r)^{1/2}}{(\rho+r)} \right\} dr. \quad (14) \end{aligned}$$

We introduce

$$x = \rho - a, \quad \xi = r - a, \quad F(\rho) \equiv h(x).$$

Then

$$\begin{aligned} h(x) + \frac{\lambda}{\pi} e^{2ika} \int_0^\infty e^{2ik\xi} \left(\frac{2a+\xi}{\xi} \right)^{1/2} \frac{1}{2a+x+\xi} h(\xi) d\xi \\ = H(x+a), \quad (x > 0). \quad (15) \end{aligned}$$

This is the required Fredholm integral equation of the second kind.

IV. DIFFRACTION BY A STRIP

Suppose that a wave $\phi_i(y, z)$ is incident on a strip lying in $z=0$, $-a \leq y \leq a$, on which $\partial\phi_i/\partial z=0$. We write the incident wave as a sum of even and odd parts:

$$\phi_i = \phi_{i1} + \phi_{i2},$$

where

$$\begin{aligned} \phi_{i1}(y, z) &= \frac{1}{2} \{ \phi_i(y, z) + \phi_i(-y, z) \}, \\ \phi_{i2}(y, z) &= \frac{1}{2} \{ \phi_i(y, z) - \phi_i(-y, z) \}. \end{aligned}$$

Similarly write the total velocity potential as $\phi_t = \phi_{t1} + \phi_{t2}$ and set

$$\phi_{t1} = \phi_{i1} + \phi_1, \quad \phi_{t2} = \phi_{i2} + \phi_2.$$

Since ϕ_1 is even about $y=0$ and tends to zero at infinity, and since $\partial\phi_1/\partial z$ is continuous across $z=0$, we can represent ϕ_1 in the form

$$\phi_1 = \left(\frac{2}{\pi} \right)^{1/2} \int_0^\infty B(t) \cos yt e^{-\gamma z} dt, \quad (z \geq 0), \quad (16a)$$

$$= - \left(\frac{2}{\pi} \right)^{1/2} \int_0^\infty B(t) \cos yt e^{\gamma z} dt, \quad (z \leq 0), \quad (16b)$$

where γ has been defined in (2). If the boundary conditions on $z=0$, namely that $\partial\phi_1/\partial z$ is known for $0 \leq y < a$ and ϕ_1 is continuous for $y > a$, are applied to (16), we obtain the dual integral equations

$$\left(\frac{2}{\pi} \right)^{1/2} \int_0^\infty \gamma B(t) \cos yt dt = g(y), \quad (0 \leq y < a), \quad (17a)$$

$$\left(\frac{2}{\pi} \right)^{1/2} \int_0^\infty B(t) \cos yt dt = 0, \quad (y > a), \quad (17b)$$

where

$$g(y) = \left\{ \partial \phi_{i1} / \partial z \right\}_{z=0}. \quad (18)$$

Suppose that (17a) equals the unknown function $g(y)$ for $y > a$. Inversion gives

$$B(t) = \left(\frac{2}{\pi} \right)^{1/2} \gamma^{-1} \int_0^\infty g(\eta) \cos \eta t d\eta. \quad (19)$$

Substitution in (17b) and interchange of orders of integration give the integral equation

$$\int_0^\infty g(\eta) \int_0^\infty \frac{1}{\gamma} \cos \eta t \cos \eta t d\eta = 0, \quad (y > a). \quad (20)$$

The kernel is of form (1) with $\nu = -\frac{1}{2}$ and we readily find, from Section II, that (20) can be written

$$\int_0^\infty g(r) \{K_0(\rho, r) + K_1(\rho, r)\} dr = 0, \quad (\rho > a), \quad (21)$$

where K_0, K_1 have been defined in (7). This is of form (10) with $\lambda = +1$. Hence (20) can be reduced to the form (15), namely

$$h(x) + \frac{e^{2ika}}{\pi} \int_0^\infty e^{2ik\xi} \left(\frac{2a + \xi}{\xi} \right)^{1/2} \frac{1}{2a + x + \xi} h(\xi) d\xi = H(x + a), \quad (x > 0), \quad (22)$$

where

$$h(x) = (\rho - a)^{1/2} e^{-ik\rho} g(\rho), \quad \rho = x + a. \quad (23)$$

The only place where k occurs explicitly under the integral sign is in the term $\exp(2ik\xi)$ and this is very convenient for asymptotic expansion provided that $h(\xi)$ has a suitable dependence on k . The important part of the integral comes from the region near $\xi = 0$ and the asymptotic expansion will depend on the behavior of $h(\xi)$ near $\xi = 0^4$ [6]. As a first approximation we have $h(x) \approx H(x + a)$. Hence, we are interested in the behavior of $H(\rho)$ for $\rho \approx a$. But this requires careful examination since the behavior of the first part of the integrand in (14) changes from $(a - r)^{1/2}$ to $(a - r)^{-1/2}$ as ρ tends to a . We should also need to be careful that $h(\xi)$ does not contain concealed exponential factors depending on k . These difficulties could be avoided by using an iterative method of solution but for a first approximation it is convenient to write

$$\int_0^\infty e^{2ik\xi} \left(\frac{2a + \xi}{\xi} \right)^{1/2} \frac{1}{2a + x + \xi} h(\xi) d\xi \approx \frac{(2a)^{1/2}}{2a + x} \int_0^\infty \xi^{-1/2} e^{2ik\xi} h(\xi) d\xi. \quad (24)$$

Then (22) gives

$$h(x) \approx H(x + a) - \frac{C}{2a + x}, \quad x > 0, \quad (25)$$

where C is a constant defined by

$$C = \pi^{-1} (2a)^{1/2} e^{2ika} \int_0^\infty \xi^{-1/2} e^{2ik\xi} h(\xi) d\xi.$$

The value of C can be determined by multiplying (25) by $x^{-1/2} \exp(2ikx)$ and integrating from zero to infinity.

$$C = \frac{\pi^{-1} (2a)^{1/2} e^{2ika} P}{1 + \pi^{-1} (2a)^{1/2} e^{2ika} Q}, \quad (26)$$

where

$$P = \int_0^\infty x^{-1/2} e^{2ikx} H(x + a) dx, \quad (27)$$

$$Q = \int_0^\infty x^{-1/2} (2a + x)^{-1} e^{2ikx} dx. \quad (28)$$

Eqs. (25) and (26) give the required approximate solution of the integral equation.

In order to determine the contribution to the far-field from the even part of the potential we note from (19) that $B(t)$ is an even function of t so that (16b) can be rewritten

$$\phi_1 = - \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^\infty B(t) e^{-iyt - \gamma|z|} dt, \quad (y \leq 0).$$

Hence, from Noble⁵ if $y = r \cos \theta$, $|z| = r \sin \theta$ (*i.e.*, θ is measured in an anti-clockwise direction from the positive y -axis), and r tends to infinity,

$$\phi_1 \sim - k^{1/2} e^{-i\pi/4} B(-k \cos \theta) \sin \theta r^{-1/2} e^{ikr}. \quad (29)$$

By substituting the value of B from (19), (23), (25) we find

$$\phi \sim - \left(\frac{2}{\pi} \right)^{1/2} \frac{e^{ikr + i\pi/4}}{(kr)^{1/2}} \left\{ \int_0^a g(r) \cos(kr \cos \theta) dr + \int_a^\infty \chi(r) \cos(kr \cos \theta) dr \right\}, \quad (30)$$

where

$$\chi(r) = (r - a)^{-1/2} e^{ikr} \{ H(r) - C(a + r)^{-1} \}. \quad (31)$$

In these equations, $g(r)$ is defined in (18) and C in (26). $H(r)$ is defined by (14) with g in place of f . Since g is an even function, this gives

$$H(r) = - \frac{1}{\pi} \int_{-a}^a e^{-iku} g(u) \frac{(a - u)^{1/2}}{(r - u)} du. \quad (32)$$

So far we have considered only the even part of the potential. The odd part can be dealt with in exactly the same way, with suitable changes in detail, *e.g.*, we must take $\lambda = -1$ instead of $+1$ in (10).

We now make a specific application of these results to diffraction of a plane wave incident normally on a strip. We shall use the following theorem,⁶ remembering the

⁴ See for instance Erdelyi [6], pp. 47-50.

⁵ Noble [16], p. 31 (1.56) and p. 35 (1.71).

⁶ See for instance Bouwkamp [2], pp. 42, 43.

distinction between diffraction by slits and strips which introduces a change of sign in this case.⁷ If, as r tends to infinity,

$$\phi \sim (2\pi/k r)^{1/2} e^{ikr + i\pi/4} A(\theta),$$

where θ is defined as in (29), then the equivalent scattering cross section is given by

$$\sigma = (2\pi/k) \operatorname{Im} A(0).$$

Hence from (30), since

$$\phi_i = e^{-ikz}, \quad g(y) = -ik, \quad (33)$$

we have

$$(2a)^{-1}\sigma = 1 - (ka)^{-1} \operatorname{Im} \int_a^\infty \chi(r) dr. \quad (34)$$

If we define

$$R = \int_0^\infty \frac{1}{x^{1/2}(2+x)} e^{2ikax} dx; \quad (35a)$$

$$S = \int_0^\infty \frac{1}{x^{1/2}(2+x)} e^{ikax} dx, \quad (35a)$$

$$T = \int_0^2 u^{1/2} e^{iku} \int_0^\infty \frac{1}{x^{1/2}(x+u)} e^{2ikax} dx, \quad (35b)$$

$$U = \int_0^2 u^{1/2} e^{iku} \int_0^\infty \frac{1}{x^{1/2}(x+u)} e^{ikax} dx, \quad (35c)$$

then (26), (31), (32), (33) give

$$\int_a^\infty \chi(r) dr = \frac{ika}{\pi} \left\{ U - \frac{\pi^{-1/2} 2^{1/2} e^{2ika} ST}{1 + \pi^{-1/2} 2^{1/2} e^{2ika} R} \right\}. \quad (36)$$

The integrals (35) can be expanded asymptotically for large k . Details will be found in [17]. The results are

$$\begin{aligned} R &\sim \frac{1}{2}\pi^{1/2} e^{i\pi/4} (2ka)^{-1/2} + 0[(ka)^{-3/2}], \\ S &\sim \frac{1}{2}\pi^{1/2} e^{i\pi/4} (ka)^{-1/2} + 0[(ka)^{-3/2}], \\ T &\sim i\pi(ka)^{-1}(\sqrt{2}-1) + \frac{1}{2}\pi^{1/2}(ka)^{-3/2} e^{2ika+i\pi/4} \\ &\quad + 0[(ka)^{-5/2}], \\ U &\sim i\pi(2ka)^{-1} + (\pi/2)^{1/2}(ka)^{-3/2} e^{2ika-i\pi/4} \\ &\quad + 0[(ka)^{-5/2}]. \end{aligned}$$

If these expansions are inserted in (36) it is found from (34) that

$$\begin{aligned} (2a)^{-1}\sigma &= 1 + \operatorname{Im} \left\{ \frac{1}{2ka} - \frac{i\sqrt{\pi}(ka)^{-3/2} e^{2ika-i\pi/4}}{1 + \frac{1}{2}(\pi ka)^{-1/2} e^{2ika+i\pi/4}} \right. \\ &\quad \left. + 0[(ka)^{-5/2}] \right\} \\ &= 1 - \frac{2\sqrt{2}}{\sqrt{\pi}} \frac{\cos(2ka - \pi/4)}{(2ka)^{3/2}} + \frac{2}{\pi} \frac{\cos 4ka}{(2ka)^2} \\ &\quad + 0[(ka)^{-5/2}]. \end{aligned}$$

This agrees with results which are already known.

⁷ *Ibid.*, pp. 38, 39.

V. DIFFRACTION BY A CIRCULAR DISK

Suppose that $\phi_i(\rho, z)$ is an axially symmetrical wave incident on a disk lying in $z=0$, $0 \leq \rho < a$ with $\partial\phi_i/\partial z = 0$ on the disk. Set $\phi_t = \phi_i + \phi$. Then ϕ can be represented in the form

$$\phi = \pm \int_{\pm}^\infty t B(t) J_0(\rho t) e^{-\gamma|z|} dt \quad (37)$$

when the upper sign holds for $z \geq 0$, the lower for $z \leq 0$. The boundary conditions on $z=0$ gives

$$\int_0^\infty \gamma t B(t) J_0(\rho t) dt = g(\rho), \quad (0 \leq \rho < a), \quad (38a)$$

$$\int_0^\infty t B(t) J_0(\rho t) dt = 0, \quad (\rho > a), \quad (38b)$$

where

$$g(\rho) = \left\{ \partial\phi_i/\partial z \right\}_{z=0}. \quad (39)$$

Suppose that the left-hand side of (38a) equals the unknown function $g(\rho)$ for $\rho > a$. Inversion yields

$$B(t) = \gamma^{-1} \int_0^\infty r g(r) J_0(rt) dr. \quad (40)$$

Substitution in (38b) and interchange of orders of integration gives the integral equation

$$\int_0^\infty r g(r) \int_0^\infty \frac{t}{\gamma} J_0(\rho t) J_0(rt) dt dr = 0, \quad (\rho > a). \quad (41)$$

The kernel is of the form (1) with $\nu=0$ and the transformation in Section II (41) shows that (41) can be written

$$\int_0^\infty r^{1/2} g(r) \{ K_0(\rho, r) - iK_1(\rho, r) \} dr = 0, \quad (\rho > a), \quad (42)$$

where $g(r)$ is known for $0 \leq r < a$. This is of form (10) with $r^{1/2}g(r)$ in place of $f(r)$ and $\lambda = -i$. In obtaining this result we have neglected higher order terms in (7a) and this is examined further in [17].

A reduction as in Section III gives an equation of type (15) which can be solved asymptotically for large k exactly as in Section IV, (22)–(28). The final result is

$$\rho^{1/2}(\rho - a)^{1/2} e^{-ik\rho} g(\rho) = H(\rho) - \frac{C}{\rho + a}, \quad (\rho > a), \quad (43)$$

where

$$\begin{aligned} H(\rho) &= -\frac{1}{\pi} \int_0^a \left\{ \frac{r^{1/2}(a-r)^{1/2}}{(\rho-r)} e^{-ikr} \right. \\ &\quad \left. - i \frac{r^{1/2}(a+r)^{1/2}}{(\rho+r)} e^{ikr} \right\} g(r) dr, \end{aligned}$$

$$C = -\frac{\pi^{-1} i (2a)^{1/2} e^{2ika} (P_1 - iP_2)}{1 - \pi^{-1} i (2a)^{1/2} e^{2ika} Q},$$

$$P_1 = -\frac{1}{\pi} \int_0^\infty x^{-1/2} e^{2ikx} \int_0^a \frac{r^{1/2}(a-r)^{1/2}}{(a+x-r)} e^{-ikr} g(r) dr dx,$$

$$P_2 = -\frac{1}{\pi} \int_0^\infty x^{-1/2} e^{2ikx} \int_0^a \frac{r^{1/2}(a+r)^{1/2}}{a+x+r} e^{ikr} g(r) dr dx,$$

and Q has been defined in (28).

As a specific application, consider diffraction of a plane wave $\phi_i = \exp(-ikz)$ normally incident on a disk. From (39)

$$g(\rho) = -ik, \quad (0 \leq \rho < a). \quad (44)$$

We use the following theorem.⁸ If as r tends to infinity

$$\phi \sim A(\theta) e^{ikr}/r, \quad (45)$$

where θ is measured from the negative z -axis, then the equivalent scattering cross section is given by

$$\sigma = (2\pi/k) \operatorname{Im} A(0). \quad (46)$$

We can find $A(0)$ from (37) since then

$$\begin{aligned} \phi(0, z) &= -\int_0^\infty t B(t) e^{-\gamma|z|} dt \\ &= -k^2 \int_0^1 B\{k(1-v^2)^{1/2}\} e^{ikv|z|} v dv \\ &\quad -k^2 \int_0^\infty B\{k(1+u^2)^{1/2}\} e^{-ku|z|} u du. \end{aligned}$$

The main contribution for large k comes from the upper limit of the first integral, and from⁹

$$\phi(0, z) \sim ikB(0) e^{ik|z|}/|z| \text{ as } z \rightarrow -\infty. \quad (47)$$

On comparing (45) and (47), we have from (46)

$$\sigma = 2\pi \operatorname{Im} \{iB(0)\}.$$

From (40) and (44) this gives

$$(\pi a^2)^{-1} \sigma = 1 - \frac{1}{ka^2} \operatorname{Im} \int_a^\infty r g(r) dr,$$

where $g(r)$ is defined in (43). If the value of $g(r)$ is inserted, the resulting integrals can be expanded asymptotically for large k . The analysis proceeds as for the corresponding equations in Section IV. The details are somewhat more complicated and can be found in [17]. The final result is

$$\begin{aligned} (\pi a^2)^{-1} \sigma &= 1 - \operatorname{Im} \left\{ -\frac{1}{ka} + \frac{2\pi^{-1/2}(ka)^{-3/2} e^{2ika - i\pi/4}}{1 + \frac{1}{2}(\pi ka)^{-1/2} e^{2ika - i\pi/4}} \right. \\ &\quad \left. + \frac{i}{4(ka)^2} + O[(ka)^{-5/2}] \right\} \end{aligned}$$

$$\begin{aligned} &= 1 - \frac{2}{\sqrt{\pi}} \frac{\sin(2ka - \pi/4)}{(ka)^{3/2}} \\ &\quad + \frac{1}{(ka)^2} \left\{ \frac{1}{\pi} \sin 2(2ka - \pi/4) - \frac{1}{4} \right\} \\ &\quad + O[(ka)^{-5/2}]. \end{aligned}$$

This agrees with the result in [13].

CONCLUSION

The main object of this paper has been to obtain the transformation of the kernel in Section II and to illustrate the application of this transformation to diffraction problems. Several other aspects are examined in [17]. In particular this reference deals with two cases not considered here, namely $\phi_i = 0$ on the strip or disk.

BIBLIOGRAPHY

- [1] B. B. Baker and E. T. Copson, "The Mathematical Theory of Huygens' Principle," Oxford University Press, London, Eng., 2nd ed.; 1950.
- [2] C. J. Bouwkamp, "Diffraction theory," *Repts. Progr. Phys.*, vol. 17, pp. 35-100; 1950.
- [3] A. P. Burger, "On the Asymptotic Solution of Wave Propagation and Oscillation Problems," Natl. Aero. Res. Inst., Amsterdam, The Netherlands, Rept. F. 157; November, 1954.
- [4] P. C. Clemmow, "Edge currents in diffraction theory," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-4, pp. 283-287; July, 1956.
- [5] W. Eckhaus, "Asymptotic solution of the two-dimensional oscillating aerofoil problem for high subsonic Mach numbers," *Proc. Ninth Intern. Congr. Appl. Mech.*, Brussels, Belgium, vol. 3, pp. 145-151; 1957.
- [6] A. Erdelyi, "Asymptotic Expansions," Dover Publications, Inc., New York, N. Y.; 1956.
- [7] E. N. Fox, "The diffraction of sound pulses by an infinitely long strip," *Phil. Roy. Trans. (London)*, vol. 241, pp. 71-103; 1948.
- [8] G. A. Grinberg, "A new method for solving problems related to the diffraction of E.M. waves by a plane with an unbounded straight slit," *J. Tech. Phys. USSR*, vol. 27, pp. 2595-2605; 1958. Translation in *Amer. Inst. Phys.*, vol. 2, pp. 2410-2419; 1959.
- [9] S. M. Karp and A. Russek, "Diffraction by a wide slit," *J. Appl. Phys.*, vol. 27, pp. 886-894; 1956.
- [10] J. B. Keller, "Diffraction by an aperture," *J. Appl. Phys.*, vol. 28, pp. 426-444; 1957.
- [11] H. Levine, "Diffraction by a Circular Aperture at High Frequencies," Inst. of Mathematical Sciences, New York University, New York, N. Y. Res. Rept. EM-84; September, 1955.
- [12] H. Levine, "Diffraction by an Infinite Slit," Applied Mathematical Statistics Lab., Stanford University, Stanford, Calif., Tech. Rept. 61; January, 1957.
- [13] H. Levine and T. T. Wu, "Diffraction by an Aperture at High Frequencies," Applied Mathematics Statistics Lab., Stanford University, Stanford, Calif., Tech. Rept. 71; December, 1957.
- [14] R. F. Millar, "The diffraction of an E.M. wave by an aperture in a plane screen," *Proc. IEE (London)*, vol. 103C, pp. 177-185; 1955; vol. 104C, pp. 87-95, 240-250, 1957.
- [15] R. F. Millar, "Diffraction by a wide slit and complementary strip," *Proc. Cambridge Phil. Soc.*, vol. 54, pp. 479-511; 1958.
- [16] B. Noble, "The Wiener-Hopf Technique," Pergamon Press, New York, N. Y.; 1958.
- [17] B. Noble, "The Integral Equations for High-Frequency Diffraction by Strips and Discs," Inst. of Mathematical Sciences, New York University, New York, N. Y., Res. Rept. EM-139, to be published.
- [18] K. Schwartschild, "Die Beugung und Polarization des Lichtes durch einen Spalt," *Math. Ann.*, vol. 55, pp. 177-247; 1902.
- [19] R. Timman, "A method for the asymptotic solution of diffraction problems," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-4, pp. 209-215; July, 1956.
- [20] N. Zitron and S. N. Karp, "Higher Order Approximation in Multiple Scattering," Inst. of Mathematical Sciences, New York University, New York, N. Y., Res. Rept. EM-126; March, 1959.

⁸ Bouwkamp [2], p. 42.

⁹ Erdelyi [6], p. 47.

Pulse Return from a Sphere*

V. H. WESTON†

Summary—The back scattering of short plane-wave harmonic pulses incident on a perfectly conducting sphere is investigated for both near and far fields. The pulse return is expressed in terms of the inverse Laplace transform of the CW back-scattered field. The inverse transform is calculated for the initial part of the pulse return using a Tauberian theorem. The latter part of the pulse return is given exactly in terms of residues representing the natural oscillations of the spheres. This residue expression converges rapidly for small ka or ka of the order of 1. However, for $ka \gg 1$, the particular residue series is slowly convergent, but the terms which are slowly convergent can again be summed using methods of contour integration to give the CW creeping waves plus transients. Calculations of the pulse return for the case $ka=1$, indicates that there is significant tail to the pulse return in the "resonance" region. For very large ka , the tail of the pulse return is the order of $1/ka$ of the head. In the high frequency limit there is no pulse distortion.

INTRODUCTION

SCATTERING of pulses by spheres has been considered by Levy and Keller,¹ Friedlander,² Wait,^{3,4} and Johler and Walters.⁵ Levy and Keller and Wait investigated the propagation around the earth of electromagnetic pulses produced by vertical electric and magnetic dipoles. Using various approaches, Friedlander investigated the reflected wave fronts and the diffracted wave fronts produced by sound waves incident on a sphere.

In this paper, the interest is in the back scattering of short plane-wave pulses by perfectly conducting spheres where the incident pulse is of time length τ and comprised of a single harmonic frequency ω . The behavior of the pulse return in the back-scattered direction, at a distance R from the center of the sphere of radius a , will be investigated. The pulse return will be a composite of the various contributions due to reflected and diffracted wave fronts, plus the fields behind these fronts.

Consideration will be made of the effect of the size of the sphere on the pulse return. The following question has to be answered. For a given fixed wavelength

of the incident pulse, what sizes of spheres will produce a return pulse with a significant tail, *i.e.*, pulse length, of back-scattered pulse much longer than that of the incident pulse?

RECEIVED PULSE

In considering the back-scattered or received pulse, only the electric field component will be considered. Since the back-scattered field for the CW case is known exactly, the relationship by which the pulse solution can be derived from the CW solution (where both satisfy the same boundary conditions on the diffracting body), by an inverse Laplace transform, will be employed.

An incident pulse of pulse length τ and frequency ω , propagating in the direction of positive z -axis, will be taken. The electric field is then given by

$$\mathbf{E}^i = \hat{\mathbf{i}}_z \begin{cases} 0 & t < z/c \\ e^{ikz-i\omega t} & z/c \leq t < z/c + \tau \\ 0 & z/c + \tau \leq t \end{cases} \quad (1)$$

which in turn can be expressed in terms of an inverse Laplace transform

$$\mathbf{E}^i = \hat{\mathbf{i}}_z \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} \frac{e^{ts} [1 - e^{-(s+i\omega)\tau}]}{(s+i\omega)} E^0 \left(\frac{is}{c} \right) ds \quad (2)$$

where $E^0(k) = e^{ikz}$.

The scattered field in the back-scattering direction is given by

$$\mathbf{E}^r = \hat{\mathbf{i}}_x \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} \frac{e^{ts} [1 - e^{-(s+i\omega)\tau}]}{(s+i\omega)} E^s \left(\frac{is}{c} \right) ds \quad (3)$$

where $E^s(k)$, the back-scattered field for the CW case,⁶ is given by

$$E^s(k) = - \sum_{n=1}^{\infty} (-1)^n \left(n + \frac{1}{2} \right) \left[\frac{h_n^{(1)}(kR) j_n(ka)}{h_n^{(1)}(ka)} + \frac{i[kR h_n^{(1)}(kR)]' [ka j_n(ka)]'}{kR [ka h_n^{(1)}(ka)]'} \right] \quad (4)$$

where the primes indicate the derivative with respect to the argument.

For $t < (R-2a)/c$, the integrand of (3) vanishes exponentially as $|s|$ approaches ∞ for $\Re(s) > 0$.⁷ Hence, one may formally extend the line integral by an integral

* The research reported in this paper has been sponsored by the Rome Air Dev. Center, Air Res. and Dev. Command, under Contract AF 30(602)-1853.

† Radiation Laboratory, University of Michigan, Ann Arbor, Mich.

¹ B. R. Levy and J. B. Keller, "Propagation of electromagnetic pulses around the earth," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-6, pp. 56-65; January, 1958.

² F. G. Friedlander, "Sound Pulses," Cambridge University Press, Cambridge; 1958.

³ J. R. Wait, "A Note on the propagation of the transient ground wave," *Proc. Symp. of VLF Radio Waves*, pp. 67-75; January, 1957. Also *Can. J. Phys.*, vol. 35, pp. 1146-1151; 1957.

⁴ J. R. Wait, "Transient fields of a vertical dipole over a homogeneous curved ground," *Can. J. Phys.*, vol. 34, pp. 27-35; January, 1956.

⁵ J. R. Johler and L. C. Walters, "Propagation of a ground wave pulse," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-7, p. 1-10; January, 1959.

⁶ J. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., Inc., New York, N. Y.; 1941.

⁷ The symbol \Re refers to the real part.

along an infinite semicircle to the right of the line $\Re(s) > b$, and equate the resulting contour integral to the sum of residues of the poles in the enclosed region. But there are no poles enclosed by the contour, since, by definition of the inverse Laplace transform, the line $\Re(s) = b$ is taken to the right of all the poles of the integrand. Thus, it can be seen that the line integral (3) is zero. Hence

$$E^r = 0 \quad t < (R - 2a)/c.$$

The local time variation of the return pulse at a particular fixed observation point, R , is to be examined. Let T be the time measured from the initial part of the return pulse:

$$t = (R - 2a)/c + T. \quad (5)$$

Define

$$F(T) = \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} \frac{\exp [T + (R - 2a)/c]s}{(s + i\omega)} E^s \left(\frac{is}{c} \right) ds. \quad (6)$$

This expression vanishes for $T < 0$. Hence

$$E^r = \hat{i}_x \begin{cases} F(T) - e^{-i\omega T} F(T - \tau) & \text{for } T \geq \tau \\ F(T) & T < \tau. \end{cases} \quad (7)$$

The remaining problem is to compute $F(T)$.

The following discussion will distinguish between two cases: (i) $0 \leq T \leq 2a/c$, and (ii) $2a/c \leq T$. An approximate method is used with case (i) and an exact method is used for case (ii).

CALCULATION OF $F(T)$ FOR $0 \leq T \leq 2a/c$

The physical significance of this time zone is the following: the contribution to the return pulse at the observation point R in the back-scattered direction is due to the reflected waves alone, without contributions from the geometric shadow region. Hence, the return pulse in this time zone can be represented by an approximation representing the reflected waves. This can be done by solving a set of "transport equations" for electromagnetic waves, as is done in Friedlander² for sound waves, or by using the Kline-Luneberg expansion, and the Tauberian theorem which states that, if

$$f(t) = \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} e^{ts} g(s) ds$$

and $g(s)$ is replaced by the asymptotic expression for large s , then the resulting transform is an asymptotic expression for $f(t)$ valid for small t .

Using the Kline-Luneberg expansion for the back-scattered field for a harmonic time-dependent field, it was shown previously that⁸

$$E^s(k) \simeq - \frac{a}{(2R - a)} e^{ik(R-2a)} \cdot \left[1 + \frac{A_1}{ka} + \frac{A_2}{(ka)^2} + \frac{A_3}{(ka)^3} + \cdots \right] \quad (8)$$

with

$$A_1 = - \frac{i2(R - a)^2}{(2R - a)^2}$$

$$A_2 = \frac{a(R - a)(2R^2 - 4Ra + 3a^2)}{(2R - a)^4}$$

$$A_3 = i \frac{139}{24 \cdot 168} + 0 \left(\frac{1}{R} \right).$$

Assume that this asymptotic expansion can be analytically continued for E^s in the s plane when ka is replaced by isa/c . Hence, for small T , one obtains

$$F(T) \simeq \frac{1}{2\pi i} \left(\frac{-a}{2R - a} \right) \int_{b-i\infty}^{b+i\infty} \frac{e^{Ts}}{(s + i\omega)} \cdot \left[1 + A_1 \left(\frac{c}{isa} \right) + A_2 \left(\frac{c}{isa} \right)^2 + \cdots \right] ds \quad (9)$$

$$\simeq - \left(\frac{a}{2R - a} \right) \left\{ e^{i\omega T} + \frac{A_1}{ka} [e^{-i\omega T} - 1] \right.$$

$$+ \frac{A_2}{(ka)^2} [e^{-i\omega T} + i\omega T - 1]$$

$$+ \left. \frac{A_3}{(ka)^3} \left[e^{-i\omega T} - 1 + i\omega T - \frac{(i\omega T)^2}{2!} \right] + \cdots \right\}. \quad (10)$$

In the above expansion it may be shown that the general term involving A_n is bounded according to:

$$\left| \frac{A_n}{(ka)^n} \left[e^{-i\omega T} - \sum_{r=0}^{n-1} \frac{(-i\omega T)^r}{(r)!} \right] \right| \leq \frac{|A_n| (cT/a)^n}{(n!)}$$

Thus, even though the Kline-Luneberg expansion may be strongly divergent, (10) could easily converge. However, to say explicitly whether it is a convergent expansion requires knowledge of the behavior of A_n for large n .

CALCULATION OF $F(T)$ FOR $T > 2a/c$

Since the integrand of (6) vanishes uniformly in an exponential manner as $|s|$ approaches infinity (real part $s < 0$), $F(T)$ may be computed exactly for all values of T greater than $2a/c$. Enclose the line integral of (6) by a contour integral to the left extending to infinity, and, using Cauchy's theorem, set the resulting closed contour integral equal to $2\pi i$ times the sum of residues enclosed, thus obtaining

$$F(T) = \text{Sum of Residues of} \quad \frac{\exp [T + (R - 2a)/c] E^s(is/c)}{(s + i\omega)}. \quad (11)$$

⁸ V. H. Weston, "Studies in Radar Cross Sections XXXIII—Exact Near-Field and Far-Field Solution for the Back Scattering of a Pulse from a Perfectly Conducting Sphere," University of Michigan, Ann Arbor, Radiation Lab. Rept. 2778-4-T, April, 1959.

The poles of the integrand are given by the zeros of $h_n^{(1)}(x)$ and $[yh_n^{(1)}(y)]'$ together with the pole at $s = -i\omega$. It is well known⁹ that for n , an integer $h_n^{(1)}(x)$ possesses n zeros which lie in the lower half x plane and lie symmetrically with respect to the imaginary axis, and that $[yh_n^{(1)}(y)]'$ possesses $(n+1)$ zeroes which lie in the lower half y plane also symmetrically with respect to the imaginary axis.

Define the zeros by the following notation:

$$\{X_n^p\} p = 1, 2, \dots, n, \text{ the zeros of } h_n^{(1)}(x) = 0. \quad (12)$$

$$\{Y_n^q\} q = 1, 2, \dots, n+1,$$

$$\text{the zeros of } [yh_n^{(1)}(y)]' = 0. \quad (13)$$

For a particular n , the zeros corresponding to $h_n^{(1)}(x)$ shall be numbered in succession $p = 1, 2, 3, \dots$, such that:

$$(2\pi - \arg X_n^1) < (2\pi - \arg X_n^2) < (2\pi - \arg X_n^3) < \dots$$

where

$$0 < \arg X_n^p < 2\pi.$$

The subscript q shall be ordered in the same manner for the zeros of $[yh_n^{(1)}(y)]'$. The zeros are discussed more fully in Appendix I (see Fig. 1).

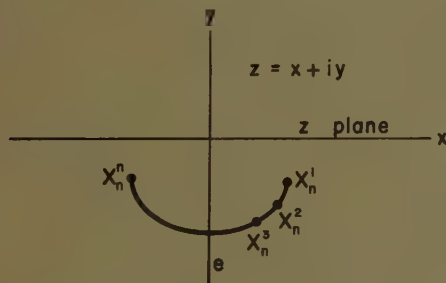


Fig. 1—Zeros of $h_n^{(1)}(x)$.

In order to simplify the analysis, set

$$A_n(X_n^p) = \frac{h_n^{(1)}(RX_n^p/a)j_n(X_n^p)}{[h_n^{(1)}(X_n^p)]'} \quad (14)$$

$$B_n(Y_n^q) = \frac{[(RY_n^q/a)h_n^{(1)}(RY_n^q/a)]'[Y_n^q j_n(Y_n^q)]'}{(RY_n^q/a)[Y_n^q h_n^{(1)}(Y_n^q)]''} \quad (15)$$

Eq. (11) can now be formally expanded in terms of the residues at the poles

$$s = -i\omega$$

$$s = -\frac{ic}{a} X_n^p, \quad n = 1, 2, \dots; p = 1, \dots, n$$

$$s = -\frac{ic}{a} Y_n^q, \quad n = 1, 2, \dots; q = 1, \dots, n+1.$$

Hence, one obtains

$$F(T) = e^{-i\omega[T+(R-2a)/c]} E^s(k) - G(T, ka) \quad (16)$$

where

$$G(T, ka) = \sum_{n=1}^{\infty} (-i)^n (n+1/2) \left\{ \sum_{p=1}^n \frac{\exp - (T + (R-2a)/c) \frac{icX_n^p}{a}}{(X_n^p - ka)} A_n(X_n^p) + i \sum_{q=1}^{n+1} \frac{\exp - \frac{ic}{a} Y_n^q (T + (R-2a)/c)}{(Y_n^q - ka)} B_n(Y_n^q) \right\}. \quad (17)$$

The expression given by (17) converges rapidly for ka less than unity, as well as for ka in the neighborhood of unity.

However, when ka increases and becomes much greater than unity the convergent is slower. Although most of the terms will decrease as ka increases because of the factors $(X_n^p - ka)^{-1}$ and $(Y_n^q - ka)^{-1}$, there exist a large number of zeros such that $(X_n^p - ka)$ and $(Y_n^q - ka)$ are the order of $(ka)^{1/3}$. These zeros are those for which the integer n is approximately equal to ka and the values of p, q are $1, 2, 3, \dots$, up to a finite number. These zeros lie in the right-half plane and can be characterized by saying that they lie in the domains

$$|X_n^p - ka| \leq (ka)^{1/2}$$

and

$$|Y_n^q - ka| \leq (ka)^{1/2}.$$

Although individually the residue terms corresponding to those zeros are large, their sum is small. It will be shown that these terms give rise to the ordinary CW creeping terms plus transient terms which behave like $1/ka$ for large ka .

To show this, consider the sum (over n) of the residues corresponding to the large zeros of $h_n^{(1)}(x)$ which have the smallest imaginary parts; i.e., the residues corresponding to the zeros X_n^1 and X_n^n .

Their contribution to $G(T, ka)$ is:

$$\sum_{n=N_0}^{\infty} (-i)^n (n+1/2) \sum_{p=1}^n \frac{A_n(X_n^p)}{(X_n^p - ka)} \cdot \exp \left[-\frac{icX_n^p}{a} (T + (R-2a)/c) \right] \{ \delta_p' + \delta_p^n \} \quad (18)$$

where $1 < N_0 < ka$

$$|ka - N_0| > (ka)^{1/2}$$

and δ_p^n is the Kronecker delta.

In order to simplify the analysis, take R to be in the far field. Near field results will be discussed below.

In the far field $A_n(X_n^p)$ becomes

$$A_n(X_n^p) = \frac{2ai(-i)^{n+1}e^{iRX_n^p/a}}{\pi R[X_n^p]^2 \{ [H_{n+1/2}^{(1)}(X_n^p)]' \}^2}. \quad (19)$$

⁹ A. Erdélyi, W. Magnus, F. Oberhettinger and F. Tricomi, "Higher Transcendental Functions," vol. 2, McGraw-Hill Book Co., Inc., New York, N. Y., 1953.

Eq. (18) can be simplified using the fact that the zeros are symmetric and thus satisfy the relation

$$X_n^{-1} = -\overline{X_n^n} \quad (20)$$

where the bar denotes the complex conjugate.

It can be shown⁸ that

$$\{[H_{n+1/2}^{(1)}(X_n^n)]'\}^2 = -\{[\overline{H_{n+1/2}^{(1)}(X_n^{-1})}']\}^2. \quad (21)$$

Hence, expression (18) becomes in the far field

$$I(ka) + \overline{I(-ka)} \quad (22)$$

where

$$I(ka) = \frac{2a}{\pi R} \sum_{n=N_0}^{\infty} \frac{(-1)^n (n+1/2) \exp[-iX_n^{-1}(Tc/a-2)]}{(X_n^{-1})^2 (X_n^{-1} - ka) \{[H_{n+1/2}^{(1)}(X_n^{-1})']\}^2}. \quad (23)$$

Since (23) is expressed in terms of the large zeros with the smallest imaginary parts, it can be shown from Weston⁸ or Franz¹⁰ that

$$I(ka) \sim \frac{-\pi a 6^{-1/3} e^{i\pi/3}}{12R[A'(\bar{t}_1)]^2} \sum_{n=N_0}^{\infty} \frac{(-1)^n (n+1/2) \exp[-iX_n^{-1}(Tc/a-2)]}{(X_n^{-1})^{2/3} (X_n^{-1} - ka)} \quad (24)$$

where $A'(\bar{t}_1)$ is the derivative of the Airy integral (see Appendix I) and

$$X_n^{-1} \sim (n+1/2) - \left(\frac{n+1/2}{6}\right)^{1/3} e^{i\pi/3} \bar{t}_1 - \frac{1}{20} \left(\frac{n+1/2}{6}\right)^{-1/3} e^{-i\pi/3} \bar{t}_1^2 \quad (25)$$

with $\bar{t}_1 = 3.372134$.

Also, the following can be obtained

$$\overline{I(-ka)} = \frac{-\pi a 6^{-1/3} e^{-i\pi/3}}{12R[A'(\bar{t}_1)]^2} \sum_{n=N_0}^{\infty} \frac{(-1)^n (n+1/2) \exp[+i\bar{X}_n^{-1}(Tc/a-2)]}{(\bar{X}_n^{-1})^{2/3} (\bar{X}_n^{-1} + ka)}. \quad (26)$$

In order to investigate the behavior of (22) for large ka , define the following functions

$$f(v) = v - \left(\frac{v}{6}\right)^{1/3} e^{i\pi/3} \bar{t}_1 - \frac{1}{20} \left(\frac{v}{6}\right)^{-1/3} e^{-i\pi/3} \bar{t}_1^2 \quad (27)$$

$$g(v) = v - \left(\frac{v}{6}\right)^{1/3} e^{-i\pi/3} \bar{t}_1 - \frac{1}{20} \left(\frac{v}{6}\right)^{-1/3} e^{+i\pi/3} \bar{t}_1^2 \quad (28)$$

and limit them to a single Riemann surface by placing the following restriction on the argument

W. Franz, "On the Green's function of the cylinder and the sphere," *Z. Naturforsch.*, vol. 9a, pp. 705-716; March, 1954.

$$-\frac{\pi}{2} \leq \arg v \leq \frac{\pi}{2}.$$

These functions have the property that

$$f(n+1/2) = X_n^{-1} \quad (29)$$

$$g(n+1/2) = \bar{X}_n^{-1}. \quad (30)$$

The series for $I(ka)$ can now be summed by equating the integral over a contour equal to the sum of residues enclosed.

Let c be a contour surrounding only the poles of $(\cos \pi v)^{-1}$ where $v = N_0 + \frac{1}{2}$, $N_0 + \frac{3}{2}$, \dots , (see Fig. 2).

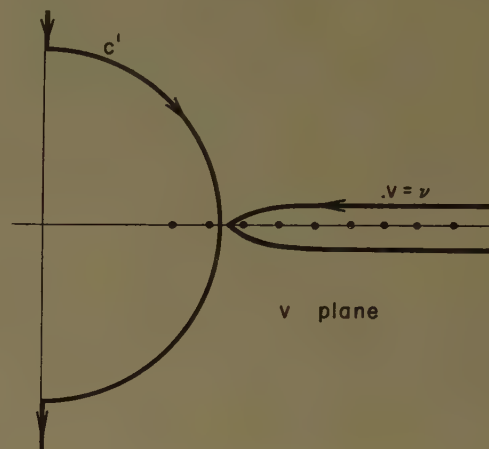


Fig. 2.

Thus

$$I(ka) \sim (-1)^m L \frac{i}{2} \int_c \frac{e^{i2m\pi v} [f(v)]^{-2/3} e^{-i[Tc/a-2]f(v)}}{\cos \pi v [f(v) - ka]} dv \quad (31)$$

where

$$L = \frac{-\pi a e^{i\pi/3} 6^{-1/3}}{12R[A'(\bar{t}_1)]^2} \quad (32)$$

and m has the values for the following time periods.

$$\begin{aligned} 2 < Tc/a < 2 + \pi & \quad \text{for } m = 0 \\ 2 + (2m-1)\pi < Tc/a < 2 + (2m+1)\pi & \quad \text{for } m \neq 0. \end{aligned} \quad (33)$$

In a similar manner

$$\overline{I(-ka)} = (-1)^m \bar{L} \frac{i}{2} \int_c \frac{e^{-i2m\pi v} [g(v)]^{-2/3}}{\cos \pi v [g(v) + ka]} \cdot e^{+i[Tc/a-2]g(v)} dv. \quad (34)$$

The integrands of both (31) and (34) will vanish in an exponential manner as $|v|$ approaches ∞ in the right half plane, provided that, given a value for T , the corresponding value of m as given by (33) is taken. The integral over the contour c in (34) may be equated to an integral over the contour c' (see Fig. 1), minus $2\pi i$ times the sum of residues enclosed by the two contours. The

contour c' is a line extending from $\infty e^{i\pi/2}$ to $N_0 e^{i\pi/2}$, plus a semicircle $|v| = N_0$ (where $-\pi/2 \leq \arg v \leq \pi/2$), and the line extending from $N_0 e^{-i\pi/2}$ to $\infty e^{-i\pi/2}$.

For the integral (31), there is a residue at $v = \nu$ where

$$f(\nu) = ka \quad (35)$$

and for the integrand (34) there is no residue enclosed. Thus, one obtains

$$\begin{aligned} I(ka) + I(-ka) &= (-1)^m L \frac{\pi \nu (ka)^{-2/3}}{\cos \pi \nu} e^{i2m\pi \nu} e^{-i\omega(T-2a/c)} \\ &+ (-1)^m L \frac{i}{2} \int_{c'} \frac{e^{i2m\pi \nu}}{\cos \pi \nu} \frac{v[f(v)]^{-2/3}}{[f(v) - ka]} e^{-if(v)[Tc/a-2]} dv \\ &+ (-1)^m L \frac{i}{2} \int_{c'} \frac{e^{-i2m\pi \nu}}{\cos \pi \nu} \frac{v[g(v)]^{-2/3}}{[g(v) + ka]} e^{+ig(v)[Tc/a-2]} dv. \quad (36) \end{aligned}$$

Since

$$\nu = ka + \left(\frac{ka}{6}\right)^{1/3} e^{i\pi/3} \bar{t}_1 + \dots, \quad (37)$$

the residue term in (36) is the CW creeping wave term corresponding to the first mode or zero of $H_p^{(1)}(ka) = 0$. The factor m has the following significance.

For $m=0$ or $2 < cT/a < 2 + \pi$ the residue term represents the contribution due to all the waves of this particular mode diffracted in the back-scattered direction. A particular wave of this mode is one which has encircled the sphere $(n + \frac{1}{2})$ times as a surface wave, then has left the surface in a direction tangent to the sphere, towards the back-scattered direction. The complete set of waves is given by $n=0, 1, 2, \dots$.

More generally, for

$$2 + (2m-1)\pi < Tc/a < 2 + (2m+1)\pi \quad (m \geq 1)$$

the residue term represents the contribution of all the waves of this particular mode which have crept around the sphere more than $(m - \frac{1}{2})$ times.

The integrals in (36) represent transients, and it can be shown that they are of order $1/ka$ for large ka . So far, only the contribution to (17) of the set of zeros (X_n', X_n'') for $n = N_0, N_0+1, \dots$ have been considered. If one considers the sets of the other zeros (X_n^2, X_n^{n-1}) , (X_n^3, X_n^{n-2}) , \dots , etc, the corresponding CW creeping waves and transient terms are obtained. A similar analysis holds for the large zeros of $[yh_n^{(1)}(y)]' = 0$ which lie near the real axes.

Since the discussion has been limited to the far field, the question remains as to what happens in the near field. The main difference between far field and near field is that the CW creeping wave terms are less in the near than in the far field. Also, the time interval given by (34) will depend on the distance R and radius a of the sphere in the near field.

For instance, the time interval will be broken up in the following forms

$$2 < Tc/a < 2 + \pi + \sin^{-1} \left(\frac{a}{R} \right) \quad m = 0$$

$$2 + \pi(2m-1) + \sin^{-1} \frac{a}{R}$$

$$< Tc/a < 2 + \pi(2m+1) + \sin^{-1} \frac{a}{R} \quad m \neq 0.$$

and in place of integral (31) one has

$$\begin{aligned} I(ka) &\sim (-1)^{m+1} \frac{LR}{2a} \int_c \frac{v[f(v)]^{1/3} e^{i(2m+1)\pi \nu}}{\cos \pi \nu [f(v) - ka]} \\ &\cdot h_p^{(1)}{}_{-1/2} \left(\frac{R}{a} f(v) \right) e^{-i[Tc/a + R/a - 2]f(v)} dv. \end{aligned}$$

PULSE RETURN

The two cases $\tau < 2a/c$ and $\tau > 2a/c$ will be considered separately. In each case, the return pulse will be divided into four time zones. The time zones and appropriate expression for E^r are given as follows.

Case (a) $\tau < 2a/c$

| Time Zone | E^r | Appropriate Expressions for $F(T)$ $F(T-\tau)$ |
|-----------------------------------|---|--|
| (i) $0 \leq T < \tau$ | $\hat{I}_x F(T)$ | (10) — |
| (ii) $\tau \leq T \leq 2a/c$ | $\hat{I}_x \{F(T) - e^{-i\omega\tau} F(T-\tau)\}$ | (10) (10) |
| (iii) $2a/c < T \leq 2a/c + \tau$ | $\hat{I}_x \{F(T) - e^{-i\omega\tau} F(T-\tau)\}$ | (16) (10) |
| (iv) $2a/c + \tau < T$ | $\hat{I}_x \{F(T) - e^{-i\omega\tau} F(T-\tau)\}$ | (16) (16) |

(38)

Case (b) $2a/c < \tau$

| Time Zone | E^r | Appropriate Expressions for $F(T)$ $F(T-\tau)$ |
|--------------------------------------|---|--|
| (i) $0 \leq T < 2a/c$ | $\hat{I}_x F(T)$ | (10) — |
| (ii) $2a/c \leq T < \tau$ | $\hat{I}_x F(T)$ | (16) — |
| (iii) $\tau \leq T \leq 2a/c + \tau$ | $\hat{I}_x \{F(T) - e^{-i\omega\tau} F(T-\tau)\}$ | (16) (10) |
| (iv) $2a/c + \tau < T$ | $\hat{I}_x \{F(T) - e^{-i\omega\tau} F(T-\tau)\}$ | (16) (16) |

(39)

Define as the "head" of the return pulse that portion corresponding to the time interval $0 \leq T < \tau$. The remainder will be defined as the "tail." The question arises as to how the pulse return varies with respect to change in size of the sphere. Consider Case (a). Neglecting the higher order terms for the present, it is seen that the head of the pulse is given by

$$E^r = -\hat{i}_x \left(\frac{a}{2R - a} \right) \cdot \left[e^{-i\omega T} + \frac{A_1}{ka} (e^{-i\omega T} - 1) + \dots \right], \quad (40)$$

and the initial part of the tail (*i.e.*, for $\tau \leq T < 2a/c$) is given by

$$E^r = -\hat{i}_x \left(\frac{a}{2R - a} \right) \left[\frac{A_1}{ka} (e^{-i\omega\tau} - 1) + \dots \right]. \quad (41)$$

Hence for $ka > 1$, the initial part of the tail is of the order A_1/ka of the head. The coefficient A_1 slowly increases as the observation point given by the coordinate R moves away from the sphere, and in the far field, A_1 approaches $-i/2$. This indicates that the magnitude of the initial part of the tail section in comparison to the magnitude of the head is largest in the far field.

The second section of the tail, *i.e.*, $2a/c < T < 2a/c + \tau$, is given by

$$E^r = \hat{i}_x \left\{ e^{-i\omega T} [e^{-ik(R-2a)} E^s(k) + \left(\frac{a}{2R - a} \right) \left(1 + \frac{A_1}{ka} + \frac{A_2}{(ka)^2} + \dots \right)] - G(T, ka) - \left(\frac{a}{2R - a} \right) e^{-i\omega\tau} \cdot \left[\frac{A_1}{ka} + \frac{A_2}{(ka)^2} (1 + i\omega\tau - i\omega T) + \dots \right] \right\}. \quad (42)$$

The first term in (42) apart from the phase factor, is

$$E^s(k) + \left(\frac{a}{2R - a} \right) \cdot e^{ik(R-2a)} \left[1 + \frac{A_1}{ka} + \frac{A_2}{(ka)^2} + \dots \right]. \quad (43)$$

This is the back-scattered field for the CW case with the reflected field component subtracted off. Hence, it represents purely the back-scattered diffracted field, *i.e.*, the creeping wave terms, plus transients.

$G(T, ka)$ may be split up into two parts. One part equals the CW creeping wave term, and the other transients of the order of $(a/R)(ka)^{-1}$. It can be shown that the CW creeping wave terms of $G(T, ka)$ cancel out those of equation (43). Hence, since the remaining terms in equation (42) are of the order of $(a/R)(ka)^{-1}$, it is seen that the section of the tail for the received pulse for

$2a/c < T < 2a/c + \tau$ is of the order of $(ka)^{-1}$ of the magnitude of the "head" of the pulse and if $\tau < \pi a/c$, and thus this section of the tail is just comprised of transients.

The remainder of the tail, *i.e.*, $T > 2a/c + \tau$, is given by

$$E^r = -\hat{i}_x [G(T, ka) - e^{-i\omega\tau} G(T - \tau, ka)]. \quad (44)$$

This is composed of the transient terms and, initially, creeping wave terms. For very large spheres and small pulses, the tail of the pulse return at most is of the order of $1/ka$ of the head of the pulse, and if $ka \gg 1$, the tail is negligible.

Now consider Case (b). For simplification assume that the pulse length τ is sufficiently greater than $2a/c$ such that the head of the pulse return has approached the CW return when $T = \tau$. This is the practical case for pulse scattering from small spheres. The head of the pulse is given by

$$(i) \quad 0 \leq T \leq 2a/c$$

$$E^r(T) = -\hat{i}_x \left[\frac{a}{2R - a} \right] \cdot \left[e^{-i\omega T} + \frac{A_1}{ka} (e^{-i\omega T} - 1) + \dots \right]. \quad (45)$$

$$(ii) \quad 2a/c < T \leq \tau$$

$$E^r(T) = \hat{i}_x [e^{-i\omega [T + (R-2a)/c]} E^s(k) - G(T, ka)]. \quad (46)$$

It is seen that initially ($T \sim 0$) the head of the pulse return is of the order of (a/R) , this contribution being due to the specular point. The pulse return builds up with contribution from the geometrically reflected waves. This is combined with the direct contribution from the shadow region and, finally, with the contribution due to the CW creeping waves. When T approaches τ , the pulse return approaches the CW return. In the resonance region, the CW return may be much less or greater than the initial part of the head of the pulse given by (45). Hence, the head of the pulse return may increase or decrease in magnitude at T approaches τ , depending upon whether the creeping wave terms are in phase with the reflected portions.

The tail of the pulse is given by

$$(i) \quad \tau < T \leq \tau + 2a/c$$

$$E^r(T) = \hat{i}_x \left\{ e^{-i\omega T} \left[e^{-ik(R-2a)} E^s(k) + \frac{a}{2R - a} \left(1 + \frac{A_1}{ka} + \frac{A_2}{(ka)^2} + \dots \right) \right] - G(T, ka) - \left(\frac{a}{2R - a} \right) e^{-i\omega\tau} \cdot \left[\frac{A_1}{ka} + \frac{A_2}{(ka)^2} (1 + i\omega\tau - i\omega T) + \dots \right] \right\}. \quad (47)$$

(ii) $\tau + 2a/c < T$

$$E^r(T) = -\hat{i}_x [G(T, ka) - e^{-i\omega\tau} G(T - \tau, ka)]. \quad (48)$$

Since the head of the pulse is assumed to have approached CW, this means that the term $G(T, ka)$ is negligible for $T \leq \tau$. Hence, the initial part of the tail given by (47) essentially behaves like the CW return with an increasing portion of the reflected wave contribution subtracted off. For large ka , this section will be of the order of $1/ka$ of the head of the pulse.

In the remaining section of the tail only the term $e^{-i\omega\tau} G(T - \tau, ka)$ has to be considered. In the time interval

$$2 < (T - \tau)c/a < 2 + \pi + \sin^{-1} \frac{a}{R},$$

contribution comes from the creeping wave terms plus transients. For large ka , this will be of the order of $(ka)^{1/3} \exp [-(ka)^{1/3} \text{ constant}]$ of the head of the pulse for the far field case. In the near field it will be much less.

For the time interval

$$2 + \pi + \sin^{-1} \frac{a}{R} < \frac{a}{R} (T - \tau)c/a < 2 + 3\pi + \sin^{-1} \frac{a}{R},$$

there will be contribution similar to that just shown, except that no longer is there any contribution due to the creeping wave terms which have crept only half-way around the sphere.

For the remaining time intervals

$$2 + (2m - 1)\pi + \sin^{-1} \frac{a}{R} < (T - \tau)c/a$$

$$< 2 + (2m + 1)\pi + \sin^{-1} \frac{a}{R} \quad (m = 1, 2, 3, \dots),$$

the CW creeping wave contribution is only from the waves which have crept at least $(m + \frac{1}{2})$ times around the sphere. This contribution is of the order of $(a/R)(ka)^{1/3} \exp \{ -(ka)^{1/3} (2m + 1) \text{ constant} \}$ in the far field, and is much less in the near field.

CALCULATED RESULT

In order to investigate the behavior in the region $ka \sim 1$, the pulse return has been calculated for Case (b); i.e., pulse length τ greater than $2a/c$. The frequency is taken such that $ka = 1$, and the distance of the observer R is given by $R/a = 200$. The initial pulse length τ is assumed to be sufficiently large such that the head of the return pulse has approached the CW return when $T = \tau - 0$, i.e.,

$$|E^r(T)| = |E^s(k)| \quad \text{for } T = \tau - 0.$$

For the case ($ka = 1$), τ should be greater than about $12a/c$.

In Fig. 3, the absolute value of the electric field for the tail of the pulse return (normalized through division by the absolute value of the CW return) is plotted versus time T measured from the initial part of the pulse return.

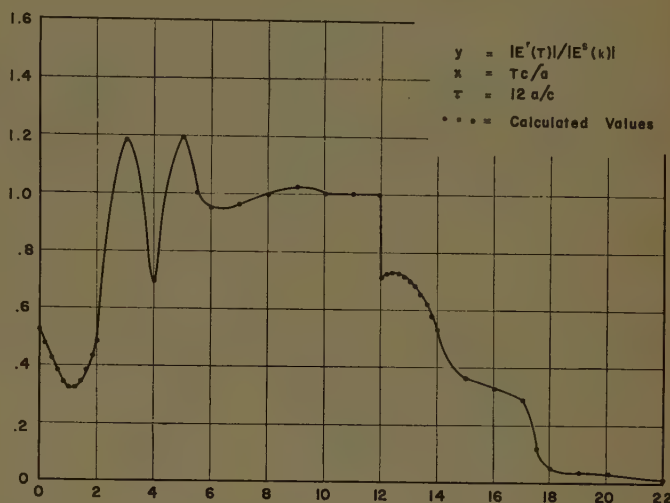


Fig. 3—Pulse return $|E^r(T)|$.

CONCLUSION

For a fixed frequency ω , the size of the sphere will affect the magnitude of the return pulse. At resonant frequencies there is a significant tail to the return pulse. At high frequencies, the tail is negligible.

For $ka > 1$, the initial part of the tail (in the time interval $\tau < T < \tau + 2a/c$) will be of the order of $1/ka$ of the head of the return pulse. The remainder of the tail ($T > \tau + 2a/c$) decays quite rapidly.

APPENDIX I

ZEROS OF $h_n^{(1)}(x)$ (n integer)

The zeros of $h_n^{(1)}(x)$ for n small have been computed, and are given in Table I. The large zeros of $h_n^{(1)}(x)$ can be estimated as follows.

Just those zeros which lie on the imaginary axis and those in the right half-plane, i.e., for real part $x \geq 0$ will be considered. The remaining zeros are obtained by symmetry.

From Watson¹¹ it can be deduced to the first approximation that the zeros satisfy the following equations.

$$\frac{n + 1/2}{x} = \cosh \gamma \quad (49)$$

$$i(n + \frac{1}{2})[\tanh \gamma - \gamma] + \frac{\pi}{4} = p\pi \quad (50)$$

¹¹ G. W. Watson, "Theory of Bessel Functions," Cambridge University Press, Cambridge, Eng., pp. 262-267; 1952.

TABLE I

| n | Zeros of $h_n^{(1)}(x)$ | | | |
|-----|----------------------------|----------------------------|-----------------------------|--------------|
| 1 | $-1.0i$ | | | |
| 2 | $\pm 0.8660254 - 1.5i$ | | | |
| 3 | $\pm 1.754381 - 1.838907i$ | $-2.322185i$ | | |
| 4 | $\pm 2.65742 - 2.10379i$ | $\pm 0.867181 - 2.8962i$ | | |
| 5 | $\pm 3.571022 - 2.324674i$ | $\pm 1.74266 - 3.35196i$ | $-3.646738i$ | |
| 6 | $\pm 4.492673 - 2.51593i$ | $\pm 2.626274 - 3.735705i$ | $\pm 0.86750965 - 4.24836i$ | |
| 7 | $\pm 5.420692 - 2.625677i$ | $\pm 3.5171 - 4.0703i$ | $\pm 1.739 - 4.758i$ | $-4.971786i$ |

TABLE II

| n | Zeros of $[Yh_n^{(1)}(Y)]'$ | | | |
|-----|-----------------------------|-----------------------------|-----------------------------|--------------------------|
| 1 | $\pm 0.8660254 - 0.5i$ | | | |
| 2 | $\pm 1.807339 - 0.7019642i$ | $-1.596072i$ | | |
| 3 | $\pm 2.757856 - 0.8428622i$ | $\pm 0.8705692 - 2.157138i$ | | |
| 4 | $\pm 3.714784 - 0.9542299i$ | $\pm 1.752303 - 2.5714i$ | $-2.948742i$ | |
| 5 | $\pm 4.676410 - 1.047674i$ | $\pm 2.644316 - 2.908062i$ | $\pm 0.8689259 - 3.554265i$ | |
| 6 | $\pm 5.641635 - 1.128905i$ | $\pm 3.54488 - 3.19524i$ | $\pm 1.74305 - 4.03354i$ | $-4.284595i$ |
| 7 | $\pm 6.609716 - 1.201203i$ | $\pm 4.45256 - 3.4476i$ | $\pm 2.6233 - 4.454i$ | $\pm 0.86840 - 4.89719i$ |

where p is an integer lying in the range

$$0 \leq p - \frac{1}{4} \leq \frac{n+1/2}{2}$$

and

$$-\frac{\pi}{2} < \arg(-i \sinh \gamma) < \frac{\pi}{2}$$

As a special case consider n an odd integer; hence, take $p = (n+1)/2$. Set $\gamma = \alpha_0 + (i\pi)/2$. Thus, (50) becomes

$$[\coth \alpha_0 - \alpha_0] - \frac{i\pi}{2} = -i \left(\frac{p-1/4}{n+1/2} \right) \pi = -\frac{i\pi}{2}.$$

Hence α_0 is a solution of the transcendental equation

$$\alpha_0 = \coth \alpha_0.$$

Thus, one obtains

$$X_{n(n+1)/2} = -i \frac{n+1/2}{\sinh \alpha_0}.$$

This is the root of $h_n^{(1)}(x)$ which is purely imaginary.

However, the interest is mainly in the roots which have the smallest imaginary part. These are given by $p = 1, 2, 3$, etc. For the case p small (50) becomes approximately

$$\tanh \gamma \sim \gamma$$

or

$$\gamma \sim 0.$$

Thus, one obtains $X \sim (n+1/2)$ and the roots with the smallest imaginary part lie in the vicinity of $|X - (n+1/2)| \sim 0(n+1/2)^{1/3}$. In this region better approximations to the zeros by using the Airy integral approximations¹⁰ can be obtained.

Let t_p be the zeros of the Airy integral $A(t)$ where

$$A(t) = \int_0^\infty \cos(x^3 - xt) dx. \quad (51)$$

Then, if X_n represent the roots of

$$H_{n+1/2}^{(1)}(X) = \sqrt{\frac{2X}{\pi}} h_n^{(1)}(X),$$

the following relationship is obtained:

$$\begin{aligned} X_n^p \sim & (n + 1/2) - \left(\frac{n + 1/2}{6}\right)^{1/3} e^{i\pi/3} \bar{t}_p \\ & - \frac{1}{20} \left(\frac{n + 1/2}{6}\right)^{-1/3} e^{-i\pi/3} \bar{t}_p^2 \end{aligned} \tag{52}$$

and for large p ,

$$\text{large } p, \bar{t}_p \sim 3 \left[\frac{\pi}{2} \left(p - \frac{1}{4} \right) \right]^{2/3}.$$

APPENDIX II

ZEROS OF $[Yh_n^{(1)}(Y)]'$

For n small, the zeros are computed and tabulated in Table II. The large zeros may be handled in the same manner as those of $h_n^{(1)}(X)$. Only the zeros with the

smallest imaginary part will be considered. These again are found from the Airy integral approximation.¹⁰ Considering only the zeros in the right half plane one obtains

$$\begin{aligned} Y_n^q = & (n + 1/2) - \left(\frac{n + 1/2}{6}\right)^{1/3} e^{i\pi/3} t_q \\ & - \left(\frac{6}{n + 1/2}\right)^{1/3} e^{-i\pi/3} \left[\frac{3}{20t_q} + \frac{t_q^3}{20} \right] \end{aligned}$$

where the t_q are the zeros of the derivative of the Airy integral $A(t)'$ and have the values for large q ,

$$t_q \sim 3 \left[\frac{\pi}{2} \left(q - \frac{3}{4} \right) \right]^{2/3}.$$

ACKNOWLEDGMENT

The author wishes to thank H. Hunter for the computational work that was performed.



Decay Exponents and Diffraction Coefficients for Surface Waves on Surfaces of Nonconstant Curvature*

JOSEPH B. KELLER† AND BERTRAM R. LEVY†

Summary—The decay exponents and diffraction coefficients for a cylindrical surface of nonconstant curvature are computed by two methods which yield the same results. The results consist of leading terms which depend upon the curvature of the surface and corrections which depend upon the derivative of the curvature. The leading terms coincide with those found previously. With these corrections, the geometric theory of diffraction can be used at longer wavelengths than before.

In the first method the solution for diffraction by an elliptic cylinder is expanded asymptotically for wavelengths small compared to the cylinder dimensions. From the expansion the decay exponents and diffraction coefficients are determined. They are then expressed in terms of the curvature and its derivative, and in this form they are assumed to apply to a cylinder of arbitrary convex cross section. This assumption is verified by comparison with the corresponding results for a parabolic cylinder. Then the same results are obtained by asymptotically solving the integral equation for the field on a cylinder of arbitrary convex cross section.

I. INTRODUCTION

WHEN a wave is incident upon an opaque object which is large compared to the incident wavelength, a shadow is formed. Some radiation penetrates into the shadow. The first quantitative analysis of this penetration effect for the case of a smooth object was made by Watson [1]. He showed that the field in the shadow of a sphere consists of a sum of modes. Each mode decays exponentially with increasing distance from the shadow boundary into the shadow. Numerous authors have pursued Watson's analysis, considering spheres which are not opaque or which are surrounded by nonuniform media. Many of these investigations are described by Bremmer [2]. Independently Franz and Depperman [3] discovered the existence of an exponentially decaying wave traveling around a circular cylinder. They also observed that this wave continues travelling into the illuminated region. These results, as well as those referred to above, pertain to bodies of constant curvature. What are the corresponding results for objects of nonconstant but slowly varying curvature?

This question was answered by the geometrical theory of diffraction introduced by Keller [4], which predicted that radiation travels along surface rays. These rays are geodesics on the surface of any object. They originate at the shadow boundary and continually shed diffracted rays which irradiate the shadow and also enter the illuminated region. A quantitative theory of the field diffracted by a cylinder of arbitrary convex cross-section was constructed with the aid of these rays

[5]. In this theory certain decay exponents and diffraction coefficients were introduced. The decay exponents determine the rate of decay of the various field modes along a surface ray. The diffraction coefficients determine the amplitudes of the various modes on a surface ray, and the amplitude of the field on the shed diffracted rays. It was assumed that the decay exponents and diffraction coefficients depend only upon local properties of the ray and the surface. By comparing the predictions of this theory with the results of Franz [6] for the circular cylinder, the decay exponents and diffraction coefficients were determined. A similar analysis was performed for three dimensional curved objects by Levy and Keller [7].

The results of the geometrical theory of diffraction have been tested by comparing them with the exact solutions of certain diffraction problems involving objects of nonconstant curvature. To make this comparison it was necessary to expand the exact solution asymptotically for wavelengths which are small compared to the dimensions of the object. This has been done for the field diffracted by a parabolic cylinder by Rice [8], an elliptic cylinder by Levy [9] and by Kazarinoff and Ritt [10], and for an ellipsoid of revolution by Levy and Keller [11] and by Kazarinoff and Ritt [12]. In all cases the leading term in the asymptotic expansion agreed precisely with the results of the geometrical theory.

We now propose to improve the geometrical theory of diffraction by an arbitrary cylinder so that it will also yield the next term in the asymptotic expansion. To this end we must determine the next terms in the expressions for the decay exponents and the diffraction coefficients. The previously determined terms involve the radius of curvature of the cylinder. The new terms will involve the derivative of the radius with respect to arc length along the cross-sectional curve. To find the new terms we shall examine the next term in the asymptotic expansion of the exact expression for the field diffracted by an elliptic cylinder. We shall express it in terms of local geometrical quantities such as the radius of curvature and its derivative. Then we shall assume that the final geometrical expression is correct for an arbitrary cylinder. As a first test of this result, we shall show that it correctly yields the next term for the field diffracted by a parabolic cylinder. Of course it also yields the correct term in the case of a circular cylinder. The same results are also obtained by asymptotically solving the integral equation for the current on the surface, as was done by Franz and Depperman [3] for the circular cylinder. These results coincide with those obtained by generalizing the results obtained for the elliptic cylinder.

The determination of these new corrections permits

* Although this paper was not presented at the Symposium, it was suggested by problems discussed there, particularly the paper by Prof. Franz which considered the same topic.

† Institute of Mathematical Sciences, New York University, New York, N. Y.

us to use our theory for longer wavelengths than could have been treated previously. The improvement resulting from the correction to the decay exponent is shown by Levy and Keller [7].

II. DIFFRACTION BY AN ELLIPTIC CYLINDER

Let us consider the field u produced by a line source parallel to the generators of an elliptic cylinder. Then u is the solution of the following problem.

$$(\Delta + k^2)u = C^{-2}\delta(\xi - \xi_0)\delta(\eta)(\cosh^2 \xi - \cos^2 \eta)^{-1} \quad (1)$$

$$\frac{\partial u(a, \eta)}{\partial \xi} = 0. \quad (2)$$

$$\lim_{r \rightarrow \infty} r^{1/2}(iku - u_r) = 0. \quad (3)$$

For simplicity the source has been taken to lie in the plane containing the major axis of the ellipse. The elliptic coordinates (ξ, η) are related to cartesian coordinates by the equations

$$x = h \cosh \xi \cos \eta, \quad (4)$$

$$y = h \sinh \xi \sin \eta. \quad (5)$$

In (4) and (5) h denotes one half the interfocal distance of the ellipses $\xi = \text{constant}$, of which $\xi = a$ is the cross-section of the cylinder.

Levy [9] has shown that on the cylinder the solution of (1), (2) and (3) can be written in the form

$$u(a, \eta) = (kh)^2 \sum_{n=1}^{\infty} b_n \frac{C_n(\eta - \pi)}{C_n'(\pi)} \frac{V_n^{(1)}(\xi_0)}{\partial V_n^{(1)'}(a)/\partial b}. \quad (6)$$

The functions C_n and $V_n^{(1)}$ are defined and asymptotically expanded for large kh by Levy [9]. A brief review of the pertinent properties of these functions follows.

The function $V_n^{(1)}$ is the outgoing solution of the equation,

$$\frac{d^2 V_n^{(1)}}{d\xi^2} - (kh)^2(b_n^2 - \cosh^2 \xi) V_n^{(1)} = 0. \quad (7)$$

For large kh it has the asymptotic expansion,

$$V_n^{(1)}(\xi) \sim \zeta^{1/4} 3^{1/3} \pi^{-1} (b_n^2 - \cosh^2 \xi)^{-1/4} A(3^{1/3} e^{-i\pi/3} (kh)^{2/3} \zeta). \quad (8)$$

Here b_n is defined by

$$V_n^{(1)'}(a) = 0. \quad (9)$$

The functions ζ and A are defined by

$$-\frac{2}{3} \zeta^{3/2} = - \int_{\cosh^{-1} b_n}^{\xi} (b_n^2 - \cosh^2 x)^{1/2} dx \quad (10)$$

$$A(t) = \int_0^{\infty} \cos(z^3 - tz) dz. \quad (11)$$

The function $C_n(\eta)$ is the even solution of

$$C_n'' + (kh)^2(b_n^2 - \cos^2 \eta) C_n = 0. \quad (12)$$

For large kh it has the asymptotic expansion,

$$C_n \sim \cos \left[kh \int_0^{\eta} (b_n^2 - \cos^2 \eta)^{1/2} d\eta \right] \{ (b_n^2 - 1)/(b_n^2 - \cos^2 \eta) \}^{1/4}. \quad (13)$$

We now specialize (6) to the case of plane wave incidence. To do this we multiply (6) by $e^{3\pi i/4} 2^{3/2} \pi^{1/2} (kh \cosh \xi_0)^{1/2} \exp[-ikh \cosh \xi_0]$ and let $\xi_0 \rightarrow \infty$. Then we obtain¹

$$u(a, \eta) = E \sum_{n=1}^{\infty} \frac{b_n C_n(\eta - \pi)}{C_n'(\pi)} \frac{\exp \left[-ikh \int_0^{\pi/2} (b_n^2 - \cos^2 \eta)^{1/2} d\eta \right]}{\partial V_n^{(1)'}(a)/\partial b}. \quad (14)$$

Here $E = e^{5\pi i/6} 2^{1/2} (kh)^{7/3}$. Upon expanding the C_n function we find that (14) becomes

$$u(a, \eta) \sim \frac{iE}{kh} \sum_{n=1}^{\infty} b_n \{ (b_n^2 - 1)(b_n^2 - \cos^2 \eta) \}^{-1/4} \frac{\exp[ikhG(\pi/2, \eta)] + \exp[ikhG(\eta, 3\pi/2)]}{\partial V_n^{(1)'}(a)/\partial b} \cdot \{ 1 - \exp[2ikhG(0, \pi)] \}^{-1}. \quad (15)$$

In (15) G is defined by

$$G(\alpha, \beta) = \int_{\alpha}^{\beta} (b_n^2 - \cos^2 \eta)^{1/2} d\eta.$$

Levy [9] computed the leading term in the asymptotic expansion of each of the summands in (15). In order to carry out this calculation it was found necessary to compute two terms in the asymptotic expansion of the eigenvalue b_n . We shall now compute a further term in the asymptotic expansion of each of the terms in (15). In order to do this we shall first compute another term in the asymptotic expansion of b_n . To do this we first observe that the leading terms in the asymptotic expansion of $V_n^{(1)'}(\xi)$ are obtained by differentiating (8). The leading term in the asymptotic expansion of $V_n^{(1)'}(\xi)$ comes from differentiating the Airy function A . Therefore, ζ_n will be nearly equal to the result obtained by Levy [9], so we write it in the form

$$\zeta_n = 3^{-1/3} e^{i\pi/3} (kh)^{-2/3} (1 + \delta_n) q_n. \quad (17)$$

Here $\zeta_n = \zeta(a)$, q_n is the n th root of the equation $A'(q_n) = 0$, and δ_n is an as yet undetermined correction which is small compared to unity.

We now set $b_n = \cosh a + \epsilon_n$ and insert this expression into (10) which determines ζ . In this way we obtain

$$\frac{2}{3} \zeta_n^{3/2} = \frac{2^{3/2}}{3} \frac{(\cosh a)^{1/2}}{\sinh a} \epsilon_n^{3/2} - \frac{2^{1/2} \epsilon_n^{5/2} (\cosh^2 a + 7)}{30 \sinh^3 a (\cosh a)^{1/2}} + O(\epsilon_n^{7/2}). \quad (18)$$

Now we insert (17) and the above form of b_n into (9) and obtain the following result for δ_n .

$$\delta_n = \frac{3^{2/3} e^{i\pi/3} (\cosh^2 a + 15 \sinh^2 a + 7)}{80 \cdot 2^{1/3} (\sinh a \cosh a)^{4/3} (kh)^{2/3}} + O((kh)^{-4/3}). \quad (19)$$

¹ The details of the evaluation of the limit are to be found in Levy and Keller [11], p. 14.

By comparing (18) and (17), and using (19) for δ_n , we determine ϵ_n . Then b_n is given by

$$b_n = \cosh a + \frac{\tau_n (\sinh a)^{2/3}}{(\cosh a)^{1/3} (kh)^{2/3}} + \frac{\tau_n^2 (\cosh^2 a + 7)}{30 (\sinh a)^{2/3} (\cosh a)^{5/3} (kh)^{4/3}} - \frac{(2 \cosh^2 a - 1)}{20 \tau_n (\sinh a)^{2/3} (\cosh a)^{5/3} (kh)^{4/3}} + O((kh)^{-2}). \quad (20)$$

In (20), the quantity τ_n is defined in terms of q_n by

$$\tau_n = \frac{q_n e^{i\pi/3}}{6^{1/3}}. \quad (21)$$

Upon substituting (20) into (16) and asymptotically expanding the integral, we obtain

$$\begin{aligned} ikhG(\alpha, \beta) = & ikh \int_{\alpha}^{\beta} (\cosh^2 a - \cos^2 \eta)^{1/2} d\eta \\ & + i(kh)^{1/3} (\sinh a \cosh a)^{2/3} \\ & \cdot \int_{\alpha}^{\beta} (\cosh^2 a - \cos^2 \eta)^{-1/2} d\eta \\ & + \frac{i\tau_n^2}{30(kh)^{1/3} (\cosh a \sinh a)^{2/3}} \\ & \cdot \int_{\alpha}^{\beta} (\cosh^2 a - \cos^2 \eta)^{-3/2} \\ & \cdot \{ (\cosh^2 a + 7)(\cosh^2 a - \cos^2 \eta) \\ & - 15 \sinh^2 a \cos^2 \eta \} d\eta \\ & - \frac{i(2 \cosh^2 a - 1)}{20 \tau_n (kh)^{1/3} (\sinh a \cosh a)^{2/3}} \\ & \cdot \int_{\alpha}^{\beta} (\cosh^2 a - \cos^2 \eta)^{-1/2} d\eta + O((kh)^{-1}). \end{aligned} \quad (22)$$

Levy [9] showed that the first two integrals in (22) have simple geometric interpretations in terms of the arc length s along the ellipse. To show this we let s_1 and s_2 be the values of s corresponding to $\eta = \alpha$ and $\eta = \beta$, respectively. Then we find that the first term on the right of (22) is just ik times the arc length

$$ik \int_{s_1}^{s_2} ds. \quad (23)$$

Similarly, the second term on the right of (22) is

$$ik^{1/3} \tau_n \int_{s_1}^{s_2} b^{-2/3} ds. \quad (24)$$

Here $b(s)$ denotes the radius of curvature of the ellipse.

We shall now express the third and fourth terms appearing on the right side of (22) in geometric terms.

The third term can be written as

$$\frac{i\tau_n^2}{30k^{1/3}} \int_{s_1}^{s_2} b^{-4/3} \left(1 + \frac{16}{9} b_s^2 - \frac{8bb_{ss}}{3} \right) ds. \quad (25)$$

The fourth term is equal to

$$- \frac{i}{20\tau_n k^{1/3}} \int_{s_1}^{s_2} b^{-4/3} \left(2 + \frac{2}{9} b_s^2 - \frac{bb_{ss}}{3} \right) ds. \quad (26)$$

The second derivatives in (25) and (26) can be eliminated by integrating by parts. Then (25) and (26), respectively, become

$$\frac{i\tau_n^2}{30k^{1/3}} \left\{ -\frac{8}{3} \frac{b_s}{b^{1/3}} \right\}_{s_1}^{s_2} + \int_{s_1}^{s_2} b^{-4/3} \left(1 + \frac{8}{9} b_s^2 \right) ds \quad (27)$$

$$- \frac{i}{20\tau_n k^{1/3}} \left\{ -\frac{1}{3} \frac{b_s}{b^{1/3}} \right\}_{s_1}^{s_2} + \int_{s_1}^{s_2} b^{-4/3} \left(2 + \frac{1}{9} b_s^2 \right) ds. \quad (28)$$

Let us now insert (22) into (15) for $u(a, \eta)$. In doing so we shall utilize the geometric forms (23), (24), (27) and (28) for the integrals in (22). We must also evaluate $\partial V_n'(a)/\partial b$ which we find, by Levy's methods [9], to be given by

$$\frac{\partial}{\partial b} V_n'(a) \sim \pi^{-1} (kh)^{4/3} 2^{1/2} (\sinh a)^{-1/2} (\cosh a)^{1/2} e^{-2i\pi/3} \cdot q_n A(q_n) + O((kh)^{2/3}). \quad (29)$$

When all these expressions are inserted into (15), we finally obtain the following asymptotic formula for u .

$$\begin{aligned} u(a, \eta) = & \frac{\pi (\cosh a)^{1/2}}{(\cosh^2 a - \cos^2 \eta)^{1/4}} \\ & \cdot \sum_{n=0}^{\infty} \{ q_n A(q_n) \}^{-1} \left(\exp \left[ik t_1 + \int_{Q_1}^P \beta_n ds \right] \right. \\ & + \exp \left[ik t_2 + \int_{Q_2}^P \beta_n ds \right] \left. \right) \{ \gamma_n + O(kh)^{-2/3} \} \\ & \cdot \left\{ 1 + \exp \left[ik T - \int_0^T \beta_n ds \right] \right\}^{-1}. \end{aligned} \quad (30)$$

In (30) the distances t_1 and t_2 and the points Q_1, Q_2 and P are as shown in Fig. 1. The distance T is the total

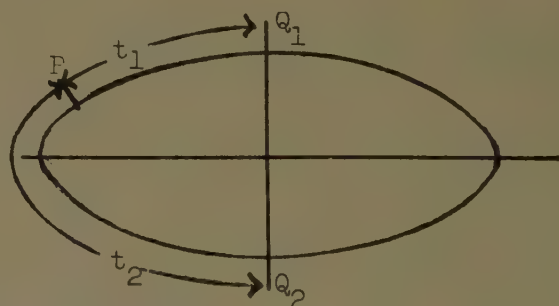


Fig. 1—A cross section of the elliptic cylinder showing the points Q_1 and Q_2 at which two incident rays are tangent to it. The incident field is a plane wave coming from the right. The tangent rays produce diffracted rays which travel distances t_1 and t_2 to a point P on the surface.

arclength of the ellipse. The quantities β_n and γ_n are defined by

$$\begin{aligned} \beta_n = & i\tau_n k^{1/3} b^{-2/3} + \frac{i\tau_n^2}{30k^{1/3}} b^{-4/3} \left(1 + \frac{8}{9} b_s^2 \right) \\ & - \frac{i}{20\tau_n k^{1/3}} b^{-4/3} \left(2 + \frac{1}{9} b_s^2 \right). \end{aligned} \quad (31)$$

$$\gamma_n = \exp \left[\frac{e^{i\pi/6} 6^{1/3} b_n(P)}{k^{1/3} b^{1/3}(P)} \left(\frac{q_n^2}{45} + \frac{1}{60 q_n} \right) \right]. \quad (32)$$

We shall now relate (30) to the geometric theory of diffraction presented by Keller [5]. When this theory is applied to the present case it yields

$$\begin{aligned} u^d(P) = 2^{1/2} 6^{1/3} \pi^{-1/2} k^{1/6} e^{-i\pi/12} b^{-1/3}(P) \sum_{n=0}^{\infty} A(q_n) B_n(P) \\ \cdot \left\{ B_n(Q_1) \exp \left[i k t_1 + \int_{Q_1}^P \beta_n ds \right] \right. \\ \left. + B_n(Q_2) \exp \left[i k t_2 + \int_{Q_2}^P \beta_n ds \right] \right\} \\ \cdot \left\{ 1 - \exp \left[i k T - \int_0^T \beta_n ds \right] \right\}^{-1}. \quad (33) \end{aligned}$$

Here β_n and B_n are, respectively, the decay exponent and diffraction coefficient of the n th mode. The leading terms β_n^0 and B_n^0 in the expressions for these quantities, as given in [5], are

$$\beta_n^0 = i \tau_n k^{1/3} b^{-2/3} \quad (34)$$

$$B_n^0(P) = \pi^{3/4} 2^{-1/4} 6^{-1/6} k^{-1/2} [q_n^{1/2} A(q_n)]^{-1} e^{i\pi/24} b^{1/6}(P). \quad (35)$$

Upon comparing (33) and (30), we see that they are identical, provided that the decay exponent β_n is given by (31) and the diffraction coefficient $B_n(P)$ by

$$B_n(P) = B_n^0(P) \gamma_n(P). \quad (36)$$

The new results, (31) and (36), for β_n and B_n agree with the previous results, (34) and (35), to the lowest order in k^{-1} . The new value of β_n is valid to $O(k^{-2/3})$ and the new value of B_n is valid to $O(k^{-3/4})$. They thus contain corrections to the previous results.

The preceding results pertain to a hard elliptic cylinder; *i.e.*, one on which $\partial u / \partial n = 0$. We have performed a similar calculation for a soft elliptic cylinder, *i.e.*, one on which $u = 0$. In this case we also find corrections to the decay exponents and to the diffraction coefficients. For the decay exponents we obtain

$$\beta_n = i \tau_n k^{1/3} b^{-2/3} + \frac{i \tau_n^2 b^{-4/3}}{30 k^{1/3}} \left(1 + \frac{8}{9} b_*^2 \right). \quad (37)$$

Here $\tau_n = 6^{-1/3} q_n e^{i\pi/3}$ and q_n is the n th root of $A(q_n) = 0$. For the diffraction coefficients we find

$$B_n(P) = B_n^0(P) \exp \left[\frac{e^{i\pi/6} 6^{1/3} b_n(P) q_n^2}{45 k^{1/3} b^{1/3}(P)} \right]. \quad (38)$$

Here $B_n^0(P)$ is the lowest order result for the diffraction coefficient, given by Keller [5], as

$$B_n^0(P) = \pi^{3/4} 2^{1/4} 6^{-2/3} k^{-1/2} [A'(q_n)]^{-1} e^{i\pi/24} b^{1/6}. \quad (39)$$

The new results, (37) and (39), contain corrections to the previous results, as in the hard case.

We now assume that results (31) and (36) apply to any hard cylinder and that (37) and (38) apply to any soft cylinder. Of course, the cylinder must have a smooth cross section. As a first check on these results we see

that when $b_* = 0$, (31) and (37) agree with results (A17a) and (A17b) of Franz [6] for a circular cylinder.

III. DIFFRACTION BY A PARABOLIC CYLINDER

As a check on the higher order corrections to the decay exponents and diffraction coefficients which were derived in Section II, we now consider the problem of diffraction by a parabolic cylinder. Our solution will closely parallel that of Rice [8]. However, we find it more convenient to use parabolic cylinder functions which differ from his, and hence we will rederive his results. We again consider the problem of evaluating the field on the surface of a hard parabolic cylinder due to an incident plane wave. For convenience we first consider the diffraction problem for an incident cylindrical wave and then obtain the plane wave result by a limiting procedure.

To formulate the diffraction problem we take the z axis of an (x, y, z) rectangular coordinate axis to be parallel to the generators of the parabolic cylinder. In the (x, y) plane we introduce parabolic coordinates (ξ, η) through

$$\begin{aligned} x &= \xi \eta \\ y &= \frac{1}{2}(\eta^2 - \xi^2). \end{aligned} \quad (40)$$

Here, $\eta > 0$ and $-\infty < \xi < \infty$. The parabolic cylinder is defined by $\eta = \text{constant} = \eta_0$. The line source is located at $y = 0$, $x = x_0$; *i.e.*, $\xi = \eta = a = x_0^{1/2}$. The wave function $u(\xi, \eta)$ then satisfies

$$u_{\xi\xi} + u_{\eta\eta} + k^2(\xi^2 + \eta^2)u = \delta(\xi - a)\delta(\eta - a). \quad (41)$$

In addition, u satisfies the boundary condition,

$$u_{\eta}(\eta_0, \xi) = 0, \quad (42)$$

and the Sommerfeld radiation condition,

$$\lim_{r \rightarrow \infty} r^{1/2}(i k u - u_r) = 0. \quad (43)$$

Now to find u we first note that the product $\phi(\xi)\psi(\eta)$ satisfies (41) with the delta functions replaced by zero if ϕ and ψ satisfy the ordinary differential equations,

$$\psi'' - k^2(b^2 - \eta^2)\psi = 0 \quad (44)$$

$$\phi'' + k^2(b^2 + \xi^2)\phi = 0. \quad (45)$$

Here b is an arbitrary separation constant. We next note that for an infinite set b_n of values of b , there exist solutions of (44), $\psi_n(\eta)$, which are "outgoing" and for which

$$\psi_n'(\eta_0) = 0.$$

We next assume that the $\psi_n(\eta)$ are complete and express u as

$$u(\xi, \eta) = \sum_{n=0}^{\infty} \Phi_n(\xi) \psi_n(\eta). \quad (46)$$

By exactly the same calculation as was carried out by Levy [9], it is easy to show that

² Since the polar coordinate variable r is equal to $\xi^2 + \eta^2/2$, we take the outgoing condition on ψ to mean that as $\eta \rightarrow \infty$, $\psi \rightarrow A e^{i k \eta^2/2}$. Here, and in the following, A will denote a generic amplitude function.

$$\int_{\eta_0}^{\infty} \psi_n(\eta) \psi_m(\eta) d\eta = -\delta_{nm} (2k^2 b_n)^{-1} \psi_n(\eta_0) \frac{\partial}{\partial b} \psi_n'(\eta_0). \quad (47)$$

Here δ_{nm} is the Kronecker delta and $\partial \psi_n'(\eta_0)/\partial b$ is the value of $\partial \psi'(\eta)/\partial b$ evaluated at $b=b_n$ and $\eta=\eta_0$. Thus, upon substituting (46) into (41), multiplying by $\psi_n(\eta)$, integrating from η_0 to ∞ , and making use of (47), we find that Φ_n satisfies

$$\Phi_n'' + k^2(b_n^2 + \xi^2)\Phi_n = \frac{-2k^2 b_n \psi_n(a)}{\psi_n(\eta_0) \partial \psi_n'(\eta_0)/\partial b} \delta(\xi - a). \quad (48)$$

To solve (48) we first characterize the solutions of (45) by means of their asymptotic expansions as $k \rightarrow \infty$. As $k \rightarrow \infty$, there exist solutions $\phi^{(1)}(\xi)$ and $\phi^{(2)}(\xi)$ having the following asymptotic expansions:

$$\phi^{(1)}(\xi) \sim (b^2 + \xi^2)^{-1/4} \exp \left[ik \int_0^\xi (b^2 + \xi^2)^{1/2} d\xi \right] \quad (49)$$

$$\phi^{(2)}(\xi) \sim (b_n^2 + \xi^2)^{-1/4} \exp \left[-ik \int_0^\xi (b^2 + \xi^2)^{1/2} d\xi \right]. \quad (50)$$

A simple calculation shows that as $|\xi| \rightarrow \infty$,

$$\phi^{(1)} \sim A \exp [ik\xi | \xi | / 2], \quad (51)$$

$$\phi^{(2)} \sim A \exp [-ik\xi | \xi | / 2]. \quad (52)$$

Thus, it is apparent that as $\xi \rightarrow \infty$, $\phi^{(1)}$ is the outgoing solution of (45), while as $\xi \rightarrow -\infty$, $\phi^{(2)}$ is the outgoing solution of (45). Since the variable ξ takes on both positive and negative values, we see that for $\xi < \xi_0$ the solution of (48) is proportional to $\phi^{(1)}(\xi)$, while for $\xi > \xi_0$ the solution of (48) is proportional to $\phi^{(2)}(\xi)$. These conditions together with the jump conditions imposed by the delta function allow a unique determination of $\Phi_n(\xi)$. We then find that for $\xi < \xi_0$,

$$u(\xi, \eta) = ik \sum_{n=0}^{\infty} b_n \psi_n(\eta) \phi_n^{(2)}(\xi) \frac{\phi_n^{(1)}(a) \psi_n(a)}{\psi_n(\eta_0) \partial \psi_n'(\eta_0)/\partial b}. \quad (53)$$

Now to pass to plane wave excitation we multiply by

$$c = e^{3\pi i/4} 2^{3/2} \pi^{1/2} k^{1/2} a e^{-ika^2} \quad (54)$$

and let $a \rightarrow \infty$. In order to evaluate this limit we require the asymptotic expansion of the function $\psi_n(a)$. Using Olver's methods [13] as in Section II we find

$$\psi_n(\eta) \sim \zeta^{1/4} 3^{1/3} \pi^{-1} (b^2 - \eta^2)^{-1/4} A(3^{1/3} e^{-i\pi/3} k^{2/3} \zeta). \quad (55)$$

Here

$$\frac{2}{3} \zeta^{3/2} = \int_b^\eta (b^2 - \eta^2)^{1/2} d\eta. \quad (56)$$

When $\eta > b$, (55) becomes

$$\psi_n(\eta) \sim e^{i\pi/12} k^{-1/6} 2^{-1} \pi^{-1/2} (\eta^2 - b^2)^{-1/4} \exp \left[ik \int_b^\eta (\eta^2 - b^2)^{1/2} d\eta \right]. \quad (57)$$

Now a simple calculation shows that as $n \rightarrow \infty$,

$$\int_b^\eta (\eta^2 - b^2)^{1/2} d\eta \sim -\frac{b^2}{2} \log \frac{2\eta}{b} + \frac{\eta^2}{2}. \quad (58)$$

Also as $\xi \rightarrow \infty$,

$$\int_0^\xi (\xi^2 + b^2)^{1/2} d\xi \sim \frac{b^2}{2} \log \frac{2\xi}{b} + \frac{\xi^2}{2}. \quad (59)$$

Upon using (58) in (57) and (59) in (49) we find

$$\lim_{a \rightarrow \infty} c \phi_n^{(1)}(a) \psi_n(a) = 2^{1/2} k^{1/3} e^{5\pi i/6}. \quad (60)$$

Thus for $\eta = \eta_0$ and for an incident plane wave, (53) becomes

$$u(\xi, \eta_0) = k^{4/3} e^{4\pi i/3} 2^{1/2} \sum_{n=0}^{\infty} b_n \frac{\phi_n^{(2)}(\xi)}{\partial \psi_n^{(1)}(\eta_0)/\partial b}. \quad (61)$$

To obtain the asymptotic expansion of (61) as $k \rightarrow \infty$ we proceed exactly as in the case of the elliptic cylinder. We first find the following three-term asymptotic expansion of b_n from the condition that $\psi_n'(\eta_0) = 0$.

$$b_n = \eta_0 + \frac{\tau_n}{k^{2/3} \eta_0^{1/3}} - \frac{7}{30} \frac{\tau_n^2}{k^{4/3} \eta_0^{5/3}} - \frac{1}{20\tau_n k^{4/3} \eta_0^{5/3}} + O(k^{-2}). \quad (62)$$

Then upon using (62) in (55) and (50) we find

$$\partial \psi_n'(\eta_0)/\partial b = \pi^{-1} e^{-2i\pi/3} q_n A(q_n) 2^{1/2} \eta_0^{1/2} k^{4/3} + O(k^{2/3}) \quad (63)$$

$$\begin{aligned} \phi_n^{(2)}(\xi) &= (\eta_0^2 + \xi^2)^{-1/4} \\ &\exp \left[-ik \int_0^\xi (\eta_0^2 + \xi^2)^{1/2} d\xi - ik^{1/3} \tau_n \eta_0^{2/3} \right. \\ &\cdot \int_0^\xi (\eta_0^2 + \xi^2)^{-1/2} d\xi - \frac{i\tau_n^2 \eta_0^{4/3}}{30k^{1/3}} \\ &\cdot \int_0^\xi (\eta_0^2 + \xi^2)^{-3/2} (8\xi^2/\eta_0^2 - 7) d\xi \\ &\left. + \frac{i}{20\tau_n k^{1/3} \eta_0^{2/3}} \int_0^\xi (\eta_0^2 + \xi^2)^{-1/2} d\xi \right]. \quad (64) \end{aligned}$$

Upon substituting (63), (64), and (62) into (61) we find

$$\begin{aligned} u(\xi, \eta_0) &= \pi \eta_0^{1/2} (\eta_0^2 + \xi^2)^{-1/4} \sum_{n=0}^{\infty} [q_n A(q_n)]^{-1} \\ &\cdot \left\{ \exp \left[-ik \int_0^\xi (\eta_0^2 + \xi^2)^{1/2} d\xi - ik^{1/3} \tau_n \eta_0^{2/3} \right. \right. \\ &\cdot \int_0^\xi (\eta_0^2 + \xi^2)^{-1/2} d\xi - \frac{i\tau_n^2 \eta_0^{4/3}}{30k^{1/3}} \\ &\cdot \int_0^\xi (\eta_0^2 + \xi^2)^{-3/2} (8\xi^2/\eta_0^2 - 7) d\xi \\ &\left. + \frac{i}{20\tau_n k^{1/3} \eta_0^{2/3}} \int_0^\xi (\eta_0^2 + \xi^2)^{-1/2} d\xi \right] \\ &\left. + O(k^{-2/3}) \right\}. \quad (65) \end{aligned}$$

In the case of the parabolic cylinder the incident rays are parallel to the x axis and the diffracted ray to the point (ξ, η_0) follows the path QP as shown in Fig. 2. A

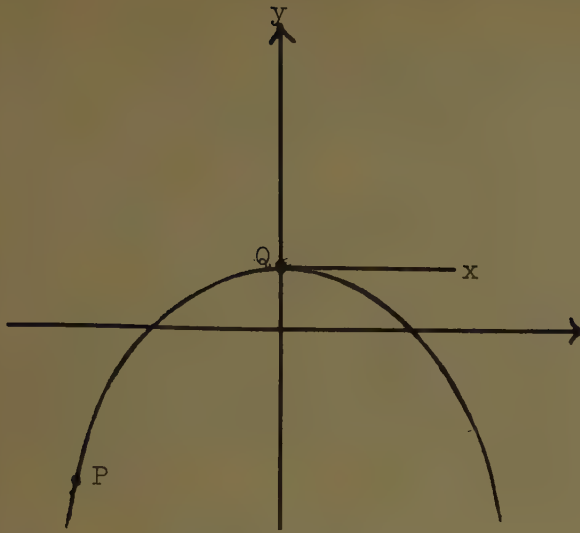


Fig. 2—A cross section of the parabolic cylinder. The incident field is a plane wave coming from the right. The diffracted ray follows the parabolic path QP .

simple calculation shows that the element of arc length along the parabola ds is given by

$$ds = (\xi^2 + \eta_0^2)^{1/2} |d\xi|. \quad (66)$$

Thus, the first term in the exponent in (65) is ikt , since ξ is negative. Similarly, we find that the radius of curvature of the parabola b is given by

$$b = \eta_0^{-1}(\eta_0^2 + \xi^2)^{3/2}. \quad (67)$$

Upon using (66) and (67), a simple calculation shows that the exponent in (65) can be written as

$$\gamma_n \exp \left[ikt + \int_Q^P \beta_n ds \right]. \quad (68)$$

Here β_n and γ_n are defined by (31) and (32) in Section II. Upon applying Keller's geometric theory [5] to the present case it is easy to show that the geometric construction agrees with (65) to lowest order in k^{-1} . Again, in (65), we have higher order corrections to the diffraction coefficients $B_n^0(P)$ and the decay exponents β_n^0 given by (34) and (35) of Section II. These corrections are identical with those given by (31) and (36) of Section II.

IV. INTEGRAL EQUATION METHOD

We will now derive the asymptotic expansion of each mode of the diffracted field on an arbitrary convex cylinder by a different method. In this method we begin with an integral equation and obtain a formal asymptotic solution of it. This asymptotic solution coincides with the expression for a mode given by the geometric theory of diffraction, with the corrected decay exponents and diffraction coefficients found in Section II. This independent derivation, which follows the procedure used by W. Franz and K. Depperman [3] in the case of a circular cylinder, confirms our previous result.

We consider the two-dimensional problem of finding a function $u(x, y)$ satisfying

$$(\nabla^2 + k^2)u = 0 \quad \text{in } D, \quad (69)$$

$$\partial u / \partial n = 0 \quad \text{on } C, \quad (70)$$

$$\lim_{r \rightarrow \infty} r^{1/2} \left(\frac{\partial u}{\partial r} - iku \right) = 0. \quad (71)$$

Here C is a given simple smooth convex curve with a piecewise continuous second derivative. If C is closed, D denotes its exterior. If C is open and extends to infinity, D denotes the nonconvex portion of the plane, bounded by C .

From (69)–(71) it follows that on C , u satisfies the following integral equation:

$$u(s) = -\frac{1}{2} \int_C u(s') \frac{\partial}{\partial n'} H_0^{(1)}[kr(s, s')] ds'. \quad (72)$$

Here s denotes arc length along C measured from some fixed point, $u(s)$ is the value of u at the point s on C , $r(s, s')$ is the distance between the points s and s' , and the normal n' points into D .

If C is closed, the only single-valued solution of (72) is $u \equiv 0$. If C is open, presumably the only bounded solution is also $u \equiv 0$. Therefore if u is to represent a mode, it must be multivalued in the former case, or unbounded in the latter case. Consequently we assume that on C a single mode u has the following asymptotic expansion for large values of k :

$$u(s) \sim \exp \left[iks + \sum_{n=1}^{\infty} v_n(s) k^{-n/3} \right]. \quad (73)$$

The coefficients $v_n(s)$ are to be determined by requiring (73) to satisfy (72) asymptotically.

Before inserting (73) into (72), we note that for large values of k the function $\partial H_0^{(1)}[kr(s, s')]/\partial n'$ has the asymptotic expansion,

$$\frac{\partial H_0^{(1)}[kr(s, s')]}{\partial n'} \sim \frac{\partial r}{\partial n'} \left(\frac{2k}{\pi r} \right)^{1/2} e^{i(kr + \pi/4)} \cdot \sum_{m=0}^{\infty} \frac{(-1)^m (0, m)}{(2ikr)^m} \left[1 - \frac{m + 1/2}{ikr} \right]. \quad (74)$$

The symbol $(0, m)$ is defined by

$$(0, m) = \Gamma(\tfrac{1}{2} + m) / m! \Gamma(\tfrac{1}{2} - m). \quad (75)$$

Now we insert (73) and (74) into (72) and then divide by the left-hand side of the resulting equation. In this way we obtain

$$1 \sim e^{-i\pi/4} (k/2\pi)^{1/2} \int_C \frac{\partial r}{\partial n'} r^{-1/2} \sum_{m=0}^{\infty} (-1)^m (0, m) (2ikr)^{-m} \cdot \left[1 - \frac{m + 1/2}{ikr} \right] \times \exp[ik(r - s - s')] \exp \left[\sum_{n=1}^{\infty} k^{-n/3} (v_n(s') - v_n(s)) \right] ds'. \quad (76)$$

In order to determine the $v_n(s)$ from (76) we first expand the integral in (76) asymptotically for large values of k . We perform this expansion by using the

concept of stationary phase. The derivative of the phase of the integrand is $1 + dr/ds'$, which vanishes if $dr/ds' = -1$. This condition is satisfied only at $s' = s$, and then only if dr/ds' denotes the one-sided derivative computed with $s' < s$. Thus to evaluate the integral we expand the integrand in the one-sided neighborhood $s' < s$ of the point $s' = s$. For this purpose we use the following expansions which are derived in the Appendix:

$$r = \sum_{n=1}^{\infty} c_n(s)(s - s')^n \quad (77)$$

$$r^{-1/2} \partial r / \partial n' = -\frac{\kappa(s)}{2} (s - s')^{1/2} \sum_{n=0}^{\infty} \rho_n(s)(s - s')^n \quad (78)$$

$$\begin{aligned} \sum_{m=0}^{\infty} (-1)^m (0, m) (2ikr)^{-m} \left[1 - \frac{m + 1/2}{ikr} \right] \\ = \sum_{n=-\infty}^{\infty} \beta_n(s, k)(s - s')^n. \end{aligned} \quad (79)$$

Here $\kappa(s)$ denotes the curvature of C . The first few of the coefficients c_n , ρ_n and β_n are listed in Table I.

TABLE I

| n | β_n | ρ_n | c_n |
|-----|-----------|---|---|
| -1 | $3i/8k$ | | |
| 0 | 1 | 1 | |
| 1 | — | $-\frac{2}{3} \frac{\dot{\kappa}}{\kappa}$ | 1 |
| 2 | — | $\frac{\ddot{\kappa}}{4\kappa} - \frac{\kappa^2}{12} + \frac{\kappa^2}{16}$ | 0 |
| 3 | — | — | $-\frac{\kappa^2}{24}$ |
| 4 | — | — | $\frac{\kappa^2}{24}$ |
| 5 | — | — | $-\frac{\dot{\kappa}^2}{90} - \frac{\kappa\ddot{\kappa}}{80} + \frac{\kappa^4}{1920}$ |

We now insert (77)–(79) into (76), making use of the explicit values of ρ_0 , c_1 , c_2 and c_3 . We also expand $v_{-1}(s')$ in a power series about the point $s' = s$. Then (76) assumes the form,

$$\begin{aligned} 1 \sim e^{3\pi i/4} \kappa(s) (k/8\pi)^{1/2} \int_{-\infty}^s (s - s')^{1/2} \\ \cdot \exp \left[-\frac{ik\kappa^2(s)}{24} (s - s')^3 \right. \\ \left. - k^{1/2} \dot{v}_{-1}(s)(s - s') \right] F(k, s, s') ds'. \end{aligned} \quad (80)$$

The function $F(k, s, s')$ appearing in (80) is defined by

$$\begin{aligned} F(k, s, s') = \exp \left[ik \sum_{n=4}^{\infty} c_n(s)(s - s')^n \right. \\ \left. + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} k^{-n/3} \frac{v_n^{(m)}(s)(s' - s)^m}{m!} \right. \\ \left. - k^{1/2} \dot{v}_{-1}(s)(s' - s) \right] \sum_{n=0}^{\infty} \rho_n(s)(s - s')^n \\ \cdot \sum_{n=-\infty}^{\infty} \beta_n(s, k)(s - s')^n. \end{aligned} \quad (81)$$

In (81) $v_n^{(m)}(s)$ denotes the m th derivative of $v_n(s)$.

To complete the asymptotic evaluation of the integral we introduce the new variable t by means of the definition,

$$s - s' = e^{-i\pi/6} \left(\frac{24}{\kappa^2(s)k} \right)^{1/3} t. \quad (82)$$

When (82) is used in (81), it shows that $F(k, s, s')$ has an expansion of the form,

$$F(k, s, s') \sim 1 + \sum_{n=1}^{\infty} k^{-n/3} b_n(t, s). \quad (83)$$

We next define $\alpha(s)$ by

$$\dot{v}_{-1}(s) = \alpha(s)(\kappa(s))^{2/3} (24)^{-1/3} e^{i\pi/6}. \quad (84)$$

Finally we insert (82)–(84) into (80), which becomes

$$1 \sim i(3/\pi)^{1/2} \int_0^{\infty} t^{1/2} e^{-\alpha t - t^3} \left(1 + \sum_{n=1}^{\infty} k^{-n/3} b_n(t, s) \right) dt. \quad (85)$$

Upon comparing coefficients of the various powers of k in the asymptotic form of the integral equation (85), we obtain the following set of equations:

$$1 - i(3/\pi)^{1/2} \int_0^{\infty} t^{1/2} e^{-\alpha t - t^3} dt = 0 \quad (86)$$

$$\int_0^{\infty} t^{1/2} e^{-\alpha t - t^3} b_n(t, s) dt = 0, \quad n = 1, 2, \dots \quad (87)$$

From these equations we shall determine the coefficients $v_n(s)$.

W. Franz [6] has shown that the left side of (86) can be rewritten in terms of the Airy function A defined in (11) of Section II. Thus (86) becomes

$$\frac{12}{\pi} e^{-i\pi/6} A' \left(-\frac{e^{i\pi/3}}{4^{1/3}} \alpha \right) A \left(-\frac{e^{-i\pi/3}}{4^{1/3}} \alpha \right) = 0. \quad (88)$$

The appropriate value of α is determined by the vanishing of the A' factor in (88). If we denote by q_n the roots of the equation $A'(q_n) = 0$, then the values α_n of α are given by

$$\alpha = \alpha_n = -e^{i\pi/3} 4^{1/3} q_n. \quad (89)$$

It will be useful to introduce the function $h(\alpha)$ defined by

$$h(\alpha) = \int_0^{\infty} t^{1/2} e^{-\alpha t - t^3} dt. \quad (90)$$

Franz [6] has shown that

$$h(\alpha) = 4^{5/6} \sqrt{3/\pi} A \left(-\frac{e^{i\pi/3}\alpha}{4^{1/3}} \right) A \left(-\frac{e^{-i\pi/3}\alpha}{4^{1/3}} \right), \quad (91)$$

and that h satisfies

$$h'''(\alpha) = -\frac{1}{6} h(\alpha) - \frac{\alpha}{3} h'(\alpha). \quad (92)$$

To determine the consequences of (87) we must first compute the b_n . We shall calculate only b_1 and b_2 . To do so we substitute (82) and (84) into (81), expand the exponential functions and multiply together the resulting series in powers of $k^{-1/3}$. In this way we obtain

$$b_1 = e^{-i\pi/6} (24)^{1/3} \frac{\ddot{\kappa}}{\kappa^{5/3}} \left\{ \left(-\frac{2}{3} - \dot{v}_0 \frac{\kappa}{\dot{\kappa}} \right) t + \frac{\alpha}{3} t^2 + t^4 \right\}, \quad (93)$$

$$b_2 = \frac{3e^{2i\pi/3} \kappa^{2/3}}{8(24)^{1/3} t} - \frac{e^{-i\pi/6} (24)^{1/3}}{\kappa^{2/3}} \dot{v}_1 t + e^{-i\pi/3} (24)^{2/3} \\ \times \left\{ \frac{t^2}{\kappa^{4/3}} \left(\frac{\ddot{\kappa}}{3\kappa} - \frac{\kappa^2}{48} + \frac{\dot{\kappa}^2}{24\kappa^2} \right) + \frac{2\alpha t^3}{3\kappa^2} \left(-\frac{\ddot{\kappa}}{6\kappa^{1/3}} - \frac{13\dot{\kappa}^2}{36\kappa^{4/3}} \right) \right. \\ + \frac{\alpha^2 \dot{\kappa}^2 t^4}{13\kappa^{10/3}} + \frac{24t^5}{\kappa^{10/3}} \left(-\frac{\kappa \ddot{\kappa}}{80} + \frac{\kappa^4}{1920} - \frac{33\dot{\kappa}^2}{720} \right) \\ \left. + \frac{\alpha \dot{\kappa}^2 t^6}{3\kappa^{10/3}} + \frac{\dot{\kappa}^2 t^8}{2\kappa^{10/3}} \right\}. \quad (94)$$

When (93) is inserted into (87), an equation for \dot{v}_0 is obtained. This equation contains integrals of the form,

$$\int_0^\infty t^{n-1/2} e^{-\alpha t - t^3} dt = (-1)^n h^{(n)}(\alpha). \quad (95)$$

In (95) the integral has been expressed in terms of the n th derivative of $h(\alpha)$ which is defined by (90). Thus from (87) we find that the right-hand side of (93) must vanish when t^n is replaced by $(-1)^{n+1} h^{(n+1)}(\alpha)$. Thus

$$\left(-\frac{2}{3} - \dot{v}_0 \frac{\kappa}{\dot{\kappa}} \right) h'' - \frac{\alpha}{3} h'''(\alpha) - h^{(V)}(\alpha) = 0. \quad (96)$$

By using (92) we find that

$$h^{(IV)}(\alpha) = -\frac{1}{2} h' - \frac{\alpha}{3} h'', \quad (97)$$

$$h^{(V)}(\alpha) = \frac{\alpha}{18} h + \frac{\alpha^2}{9} h' - \frac{5}{6} h''. \quad (98)$$

When (92) and (98) are used in (96), the following expression for \dot{v}_0 results:

$$\dot{v}_0 = \frac{1}{6} \frac{\dot{\kappa}}{\kappa}. \quad (99)$$

Upon integrating (99) we finally obtain for v_0 the expression,

$$v_0 = \log \kappa^{1/6} + \delta. \quad (100)$$

Here δ is an integration constant.

The analysis of (87) for the case $n=2$ proceeds in exactly the same way. In this case we obtain the condition that the right-hand side of (94) vanishes when t^n

is replaced by $(-1)^{n+1} h^{(n+1)}(\alpha)$.³ In order to simplify the resulting expression we must express the sixth through ninth derivatives of h in terms of h , h' , and h'' . Upon doing this we find

$$h^{(VI)} = \frac{7}{36} h + \frac{5\alpha}{9} h' + \frac{\alpha^2}{9} h'', \quad (101)$$

$$h^{(VII)} = -\frac{\alpha^2}{54} h + \frac{3}{4} h' - \frac{\alpha^3}{27} h'' + \frac{7}{9} \alpha h''', \quad (102)$$

$$h^{(VIII)} = -\frac{\alpha}{6} h - \frac{7}{18} \alpha^2 h' + \frac{55}{36} h'' - \frac{\alpha^3}{27} h''', \quad (103)$$

$$h^{(IX)} = \frac{\alpha^3}{162} h - \frac{91}{216} h' - \frac{157}{108} \alpha h'' + \frac{\alpha^4}{81} h''' - \frac{\alpha^2}{2} h'''. \quad (104)$$

We also note that

$$h''(\alpha) = -\frac{\alpha}{6} h(\alpha). \quad (105)$$

This result follows upon differentiating (91) twice, then using the equation satisfied by $A(x)$,

$$A'' + \frac{x}{3} A = 0, \quad (106)$$

and finally noting that when α is defined by (89),

$$A' \left(-\frac{e^{i\pi/3}\alpha}{4^{1/3}} \right) = 0. \quad (107)$$

We now insert the preceding relations for the derivatives of h , together with (99) and (105) into b_2^* . We then find that the coefficient of h' vanishes and that the equation $b_2^* = 0$ may be solved to yield

$$\dot{v}_1 = 6(24)^{1/3} \kappa^{2/3} \alpha^{-1} e^{-i\pi/6} \\ \times \left\{ \frac{\kappa^{2/3}}{64} - \frac{1}{6\kappa^{4/3}} \left(\frac{1}{3} \frac{\ddot{\kappa}}{\kappa} - \frac{1}{48} \kappa^2 + \frac{1}{24} \frac{\dot{\kappa}^2}{\kappa^2} \right) \right. \\ + \frac{\alpha^3}{27\kappa^2} \left(\frac{\ddot{\kappa}}{6\kappa^{1/3}} + \frac{13}{36} \frac{\dot{\kappa}^2}{\kappa^{4/3}} \right) \\ + \frac{7\alpha^3}{648} \frac{\dot{\kappa}^2}{\kappa^{10/3}} + \frac{24}{\kappa^{10/3}} \left(\frac{7}{36} - \frac{\alpha^3}{54} \right) \\ \times \left(\frac{\kappa \ddot{\kappa}}{80} - \frac{\kappa^4}{1920} + \frac{33}{720} \dot{\kappa}^2 \right) - \frac{4\alpha^3}{81} \frac{\dot{\kappa}^2}{\kappa^{10/3}} \\ \left. + \frac{1}{2} \frac{\dot{\kappa}^2}{\kappa^{10/3}} \left(\frac{29\alpha^3}{324} - \frac{91}{216} \right) \right\}. \quad (108)$$

We next make use of (89) of this Section and (21) of Section II and set $\kappa^{-1} = b$ = the radius of curvature of C . Then a straightforward calculation shows that (108) may be written as

$$\dot{v}_1 = \frac{i\tau_n^2}{30} \left(b^{-4/3} + \frac{16}{9} b^{-4/3} b_s^2 - \frac{8b^{-1/3}}{3} b_{ss} \right) \\ - \frac{i}{20\tau_n} \left(2b^{-4/3} + \frac{2}{9} b^{-4/3} b_s^2 - \frac{b^{-1/3}}{3} b_{ss} \right). \quad (109)$$

³ In order to avoid writing cumbersome equations we denote by b_2^* the right-hand side of (94) with t^n replaced by $(-1)^{n+1} h^{(n+1)}(\alpha)$.

Let us now combine our results (84), (89), (100), and (109). By inserting them into (73) we obtain the asymptotic expansion of $u(s)$ on the cylinder C up to and including terms in $k^{-1/3}$.

$$u \sim E b^{-1/6}(s) \exp \left[i k s + i k^{1/3} \tau_n \int^s b^{-2/3}(s) ds + \frac{i k^{-1/3}}{30} \tau_n^2 \int^s \left(b^{-4/3} + \frac{16}{9} b^{-4/3} b_s^2 - \frac{8 b^{-1/3}}{3} b_{ss} \right) ds - \frac{i k^{-1/3}}{20 \tau_n} \int^s \left(2 b^{-4/3} + \frac{2}{9} b^{-4/3} b_s^2 - \frac{b^{-1/3}}{3} b_{ss} \right) ds + \dots \right]. \quad (110)$$

Here E denotes an arbitrary constant.

Let us now compare the result (110) with the expression for a mode given by the geometric theory of diffraction [5]. That theory yields for u a single term of the sum in (33) of Section II. Let us insert into that equation the improved decay exponents (99) and diffraction coefficients (104). Then we find that each term of (101) coincides with (110) provided that the product of the constant coefficients in (101) is equated to the constant E in (110). This agreement between the two methods of obtaining the improved decay exponents and diffraction coefficients again confirms the results of Section II. The method of the present section can also be modified to apply to soft cylinders, on which $u = 0$.

V. APPENDIX

In order to calculate the quantity $r(s, s')$ in the neighborhood of $s = s'$ we first observe that if $\mathbf{x}(s)$ is the position vector to the curve C , then

$$r^2 = (\mathbf{x}(s) - \mathbf{x}(s'))^2. \quad (111)$$

By Taylor's theorem,

$$\mathbf{x}(s') - \mathbf{x}(s) = \sum_{n=1}^{\infty} (s' - s)^n \frac{\mathbf{x}^{(n)}(s)}{n!}. \quad (112)$$

Thus, upon taking the dot product of (112) with itself we find

$$r^2 = \sum_{\rho=2}^{\infty} (s' - s)^{\rho} b_{\rho}. \quad (113)$$

Here

$$b_{\rho} = \sum_{n=1}^{\rho-1} \frac{\mathbf{x}^{(n)}(s) \cdot \mathbf{x}^{(\rho-n)}(s)}{n!(\rho-n)!}. \quad (114)$$

Since s is arc length along C , we have

$$\dot{\mathbf{x}}(s) \cdot \dot{\mathbf{x}}(s) = 1, \quad (115)$$

and

$$\ddot{\mathbf{x}}(s) \cdot \dot{\mathbf{x}}(s) = 0. \quad (116)$$

From the Frenet equations of differential geometry we have

$$\ddot{\mathbf{x}}(s) = \kappa \kappa^{-1} \dot{\mathbf{x}} - \kappa^2 \dot{\mathbf{x}}. \quad (117)$$

Upon using (115), (116), and (117) recursively to obtain

the higher derivatives of $\mathbf{x}(s)$ in terms of $\mathbf{x}(s)$ and $\dot{\mathbf{x}}(s)$ we see that b_{ρ} can be expressed in terms of κ and its derivatives. In this way we find

$$b_2 = 1; b_3 = 0; b_4 = \frac{-\kappa^2}{12}; b_5 = \frac{-\kappa \dot{\kappa}}{12}; b_6 = \frac{-1}{45} \dot{\kappa}^2 - \frac{1}{40} \kappa \ddot{\kappa} + \frac{\kappa^4}{360}. \quad (118)$$

Then upon taking the square root of the right-hand side of (113) we find

$$r = \sum_{n=1}^{\infty} c_n (s - s')^n. \quad (119)$$

Here

$$c_1 = b_2; c_2 = b_3; c_3 = \frac{b_4}{2}; c_4 = \frac{-b_5}{2}; c_5 = \left(\frac{b_6}{2} - \frac{b_4^2}{8} \right). \quad (120)$$

From (118) and (120) the entries for c_n in Table I are obtained.

In order to calculate $r^{-1/2} \partial r / \partial n'$ we note that the unit normal to C at s' is $\kappa^{-1}(s') \ddot{\mathbf{x}}(s') = \mathbf{v}_2(s')$ and, hence,

$$\frac{\partial r}{\partial n} = \nabla(r) \cdot \mathbf{v}_2(s') = \frac{\mathbf{r} \cdot \mathbf{v}_2(s')}{r}. \quad (121)$$

Thus by making use of (112) and the Taylor expansion of $\mathbf{v}_2(s')$ about $s = s'$, we obtain

$$r \frac{\partial r}{\partial n'} = \sum_{\rho=1}^{\infty} (s' - s)^{\rho} f_{\rho}. \quad (122)$$

Here,

$$f_{\rho} = \sum_{k=1}^{\rho} \frac{\mathbf{x}^{(k)}(s) \cdot \mathbf{v}_2^{(\rho-k)}(s)}{k!(\rho-k)!}. \quad (123)$$

Again upon using the Frenet equations recursively, the coefficients f_{ρ} may be easily evaluated to obtain

$$f_1 = 0; f_2 = -\frac{\kappa}{2}; f_3 = -\frac{\dot{\kappa}}{3}; f_4 = -\frac{\ddot{\kappa}}{8} + \frac{\kappa^3}{24}. \quad (124)$$

Now by applying the binomial theorem to (113) we find

$$r^{-3/2} = (s - s')^{-3/2} \left[b_2 - \frac{3}{4} b_4 (s - s')^2 + \frac{3}{4} b_6 (s - s')^4 + \dots \right]. \quad (125)$$

Thus upon multiplying (125) and (122), we find

$$r^{-1/2} \frac{\partial r}{\partial n'} = -\frac{\kappa(s)}{2} (s - s')^{1/2} \sum_{n=0}^{\infty} \rho_n(s) (s - s')^n, \quad (126)$$

where

$$\rho_0 = 1; \rho_1 = \frac{2f_3}{\kappa}; \rho_2 = -\frac{2f_4}{\kappa} - \frac{3}{4} b_4. \quad (127)$$

Thus from (118), (124), and (127) we obtain the values of ρ_n given in Table I. Upon using (119) and the values of $(0, m)$ given in Section IV, it is a simple matter to calculate the values of β_0 and β_{-1} as given in Table I and to conclude that $\beta_n(s - s')^n = o(k^{-2/3})$ for $n \neq 0, -1$ and $s - s' = O(k^{-1/3})$.

REFERENCES

- [1] G. N. Watson, "The diffraction of electric waves by the earth and the transmission of electric waves around the earth," *Proc. Roy. Soc. (London)*, vol. A95, pp. 83-99, October, 1918; pp. 546-563, July, 1919.
- [2] H. Bremmer, "Terrestrial Radio Waves," Elsevier Publishing Co., New York, N. Y.; 1949.
- [3] W. Franz and K. Depperman, "Theorie der Beugung am Zylinder unter Berücksichtigung der Kriechwelle," *Ann. der Phys.*, vol. 10, pp. 361-373; 1952.
- [4] J. B. Keller, "A geometric theory of diffraction," in "Calculus of Variations and its Applications," Proceedings of Symposia in Applied Mathematics, McGraw-Hill Book Co., Inc., New York, N. Y.; vol. 8, pp. 27-52; 1958.
- [5] J. B. Keller, "Diffraction by a convex cylinder," *IRE TRANSMISSION ON ANTENNAS AND PROPAGATION*, vol. AP-4, pp. 312-321; July, 1956.
- [6] W. Franz, "Über die Greenschen Funktionen des Zylinders und der Kugel," *Z. Naturforsch.*, vol. 9a, pp. 705-716; September, 1954.
- [7] B. R. Levy and J. B. Keller, "Diffraction by a smooth object,"

- Commun. Pure and Appl. Math.*, vol. 12, pp. 159-209; February, 1959.
- [8] S. O. Rice, "Diffraction of plane radio waves by a parabolic cylinder," *Bell Syst. Tech. J.*, vol. 33, pp. 417-502; March, 1954.
- [9] B. R. Levy, "Diffraction by an Elliptic Cylinder," *Inst. of Math. Sci., Div. of EM Res., New York University, New York, N. Y.*, Res. Rept. No. EM-121, 1958; and *J. Math. and Tech.*, March, 1960.
- [10] N. D. Kazarinoff and R. K. Ritt, "Scalar Diffraction by an Elliptic Cylinder," *Res. Inst., University of Michigan, Ann Arbor, Mich.*, Scientific Rept. No. 2871-2-T; June, 1959.
- [11] B. R. Levy and J. B. Keller, "Diffraction by a Spheroid," *Inst. of Math. Sci., Div. of EM Res., New York University, New York, N. Y.*, Res. Rept. No. EM-130, 1959; and *Canad. J. Phys.*, January, 1960.
- [12] R. K. Ritt and N. D. Kazarinoff, "Studies in Radar Cross-Sections XXX, The Theory of Scalar Diffraction with Application to the Prolate Spheroid," *University of Michigan Res. Inst., Scientific Rept. No. 4*, 1958; and *Ann. Phys.*, vol. 6, p. 227, 1959.
- [13] F. W. J. Olver, "The asymptotic solution of linear differential equations of the second order for large values of a parameter," *Phil. Trans. Roy. Soc. London, Series A*, vol. 247, pp. 307-368; 1954.

New Results in Backscattering from Cones and Spheroids*

A. OLTE† AND S. SILVER‡

Summary—Experimental results are given for backscattering cross sections of cones, prolate spheroids, and oblate spheroids. A brief resume is given first of the image-plane technique used in measuring cross sections. Calibration data show that for cross sections of the order of spheres of 3-inch diameter the attainable accuracy is better than 0.5 db. Cross-section data for various aspects of the targets are compared with geometrical optics cross sections and those obtained by various other theoretical procedures. Some interpretation of the observed phenomena is made in terms of high-frequency approximation techniques.

INTRODUCTION

THE scattering of electromagnetic waves by figures of revolution continues to engage considerable attention as is evidenced by the program of this symposium and of a number of recent meetings. This paper is a presentation of the results of an experimental investigation of backscattering (or radar) cross sections of right circular cones, prolate spheroids, and oblate spheroids, in each case for the aspects for which the direction of propagation of the incident wave coincides with a principal axis of the figure. The backscattering cross sections of these figures have been the subject of considerable theoretical study recently and we hope that our data will provide not only a test of the current state of theoretical developments but will also be a guide to future work.

The families of figures which we have studied have interesting interrelationships. Both the cones and prolate spheroids comprise a class of figures having relatively large length to transverse diameter ratios. The cones have two singularities, the tip and the edge of the base, the latter forming the geometrical shadow boundary for the aspects considered in this paper. The prolate spheroids have continuous curvature, and corresponding to that the scattering is of a different order of magnitude. The oblate spheroids form a progressive sequence from a sphere to a disc and we have a range of dimensions from cases when all dimensions are large compared with wavelength, to mixed cases when some dimensions are large and some small compared with wavelength. Data are given for the several possible orientations of the polarization of the incident wave relative to the figure axis.

It will be recalled that at the last General Assembly of URSI Siegel¹ gave a review of the theoretical work which had been done on scattering, and discussed in particular the limiting cases when the wavelength is very long and when the wavelength is very short compared with the dimensions of the scatterer. Recently, Keller² has published a report on the application of his geometrical theory of diffraction to cones, another way

* This work was supported by the Office of Naval Research, under Contract N7-onr-29529.

† Dept. Elec. Engrg. and Radiation Lab., University of Michigan, Ann Arbor, Mich. Formerly Dept. Elec. Engrg., Electronics Res. Lab., University of California, Berkeley, Calif.

‡ Dept. Elec. Engrg., Electronics Res. Lab., University of California, Berkeley, Calif.

¹ K. M. Siegel, "Far field scattering from bodies of revolution," *J. Appl. Sci. Res.*, sect. B, vol. 7, pp. 292-328; 1958. See also, K. M. Siegel, R. Goodrich, and V. Weston, "Comments on far-field scattering from bodies of revolution," *J. Appl. Sci. Res.*, sect. B, vol. 8; 1959.

² J. Keller, "Backscattering from a finite cone," *New York University, Inst. of Math. Sciences, Div. Electromag. Res. Rept. No. EM-127*; February, 1959.

of treating the high frequency case. The work on scattering is in a quite satisfactory form for the two limiting cases even though there are differences in detail between various approaches and the quantitative tests of the theories are far from complete. The region between the asymptotic limits remains to be dealt with, and here experimental data should be very helpful.

Directed and systematic experimental investigations have been actually fewer in number than the theoretical studies. Honda, Silver, and Clapp^{3,4} developed an image-plane type of scattering range for measuring backscattering cross sections with a high degree of accuracy and for measuring differential cross sections over a complete 360° in the symmetry planes of the scatterer. They obtained some data on cones and prolate spheroids. Keys and Primich⁵ made an extensive study of backscattering from a series of cones passing from long thin cones to disks. Brysk, Hiatt, Weston and Siegel⁶ have reported recently measurements on cones made at the University of Michigan and give in that report additional discussion of their theoretical work. Angelakos and Shostak⁷ made a series of measurements of backscattering from cones as a function of aspect and showed the large specular reflection lobes and also the diffraction lobes that appear. Our present results added to those of Keys and Primich and those of the University of Michigan group make a fairly comprehensive body of data on backscattering from cones.

EXPERIMENTAL TECHNIQUES

The half-space technique and procedures which we have used are essentially the same as those discussed earlier by Honda, Silver, and Clapp.⁴ A number of improvements have been made in the circuitry and in the environmental conditions. Olte⁸ has made a detailed study of the technique and the sources of error in the measurements and will publish a summary of his work in a separate paper in the near future. Here we shall give only a brief description of the system and some calibration data to show the precision and accuracy which we can attain.

The image plane and some of the associated gear are shown in Fig. 1. The platform which can be seen extending over the plane allows the experimenter to align the

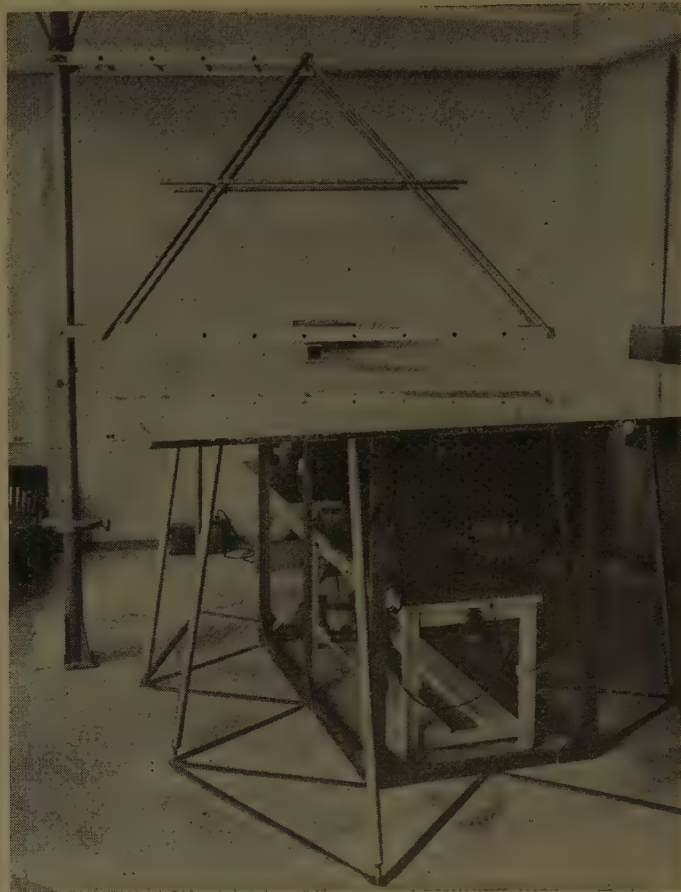


Fig. 1—Photograph of the scattering range.

target with the horn; it swings away out of the path of the incident wave. The essential elements of the circuit and measurement procedure are shown by the block diagram of Fig. 2. The horn constitutes one arm of a magic-T and is used for transmitting and receiving. The backscattered wave appears in the system as a mismatch of the bridge circuit and hence as an output from the *E*-plane arm. The return signal is phase-modulated by a rotating phase shifter so that heterodyne detection can be used. The beat frequency, which is in this case an audio frequency, is used directly so that second detection is not needed. This technique is an adaptation of a system employed by D. H. Ring and others in the Bell Telephone Research Laboratories.

The target is moved slowly away from the horn along the center line of the horn and the backscattering is measured as a function of distance. For each run of the target a corresponding run is made with a reference sphere. This is done for three reasons: first, we must eliminate the contribution of background reflections coming from the edges of the ground plane, the walls, and from residual mismatch in the bridge. The return from the target oscillates with distance as the target and background reflections come in and out of phase. The interactions involved are more complicated, actually, but the basic aspects are the same. It is possible then to determine the target return by drawing the mean curve through the measured one. Second, it is necessary to

³ J. Honda, "Scattering of microwaves by figures of revolution," Ph.D. thesis, University of California, Berkeley, 1956. See also, 1957 WESCON CONVENTION RECORD, pt. I, pp. 151-157.

⁴ J. Honda, S. Silver, and F. D. Clapp, University of California, Berkeley, Electronics Res. Lab. Rept. No. 84, Issue 232; March 16, 1959.

⁵ J. E. Keys and R. I. Primich, "The nose-on radar cross sections of conducting right circular cones," *Canad. J. Phys.*, vol. 37, pp. 521-522; April, 1959.

⁶ H. Brysk, R. E. Hiatt, V. H. Weston, and K. M. Siegel, "The Nose-On Radar Cross Sections of Finite Cones," Radiation Lab., University of Michigan, Ann Arbor, Mich., *Canad. J. Phys.*, vol. 37, pp. 675-679; May, 1959.

⁷ D. J. Angelakos and A. Shostak, "Back-Scatter From a Right Circular Cone," University of California, Berkeley, Electronics Res. Lab. Rept. No. 70, Issue No. 191; July 26, 1957.

⁸ A. Olte, "Precision Measurement of Scattering from Figures of Revolution," Ph.D. Dissertation, University of California, Berkeley; 1959.

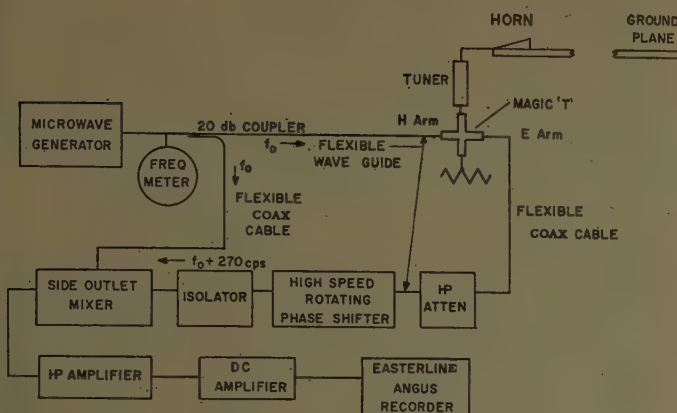


Fig. 2—Block diagram of the measuring system.

eliminate the constants of the system in order to obtain the cross section of the measured output from the bridge. By getting the corresponding return from a sphere, we can establish the ratio of the return from the target to that from the sphere for which the cross section can be computed from the exact theory. Third, it is necessary to ascertain the limiting value of the return as the distance goes to infinity in order to determine the true cross section. When the horn is in the far zone of the target, the return should vary inversely as the square of the distance and the ratio of the target return to that from the reference sphere should be independent of distance. A plot of the ratio of target return to sphere return thus enables us to establish the far zone limiting value and from that the ratio of the cross section of the target to that of a reference sphere.

The range of the system and the accuracy of the cross sections are determined by the residual null of the bridge and the total power available for illuminating the scatterer. The far zone of the horn which we are using begins at $2d^2/\lambda = 25$ inches at the nominal frequency of 9300 mc. If the object is of such size that we must start the run at a distance of 50 inches from the aperture, the minimum signal we can measure is 45 db below that from a sphere of 6-inch diameter. If the objects are smaller, we can start closer to the horn and, taking the limiting case when we can start at 25 inches from the horn, the smallest signal we can measure is 50 db below that from a 6-inch diameter sphere.

In general we use two reference spheres as standards, one having a diameter of 6 inches and the other a diameter of 3 inches. We have actually a range of spheres which we have measured with respect to one another and we can, therefore, choose for a given target a reference sphere whose return is as nearly the same in magnitude as possible. The measurements of the cross sections of the sphere calibrate the system for us and give us a direct evaluation of the attainable accuracy. Table I shows typical results. The calculated values have been computed from the harmonic series solution for the scattered field. The evaluation of the measurement technique does not, of course, include errors in the shape

TABLE I
CALIBRATION DATA σ/σ_{ref} IN DB; REFERENCE
SPHERE DIAMETER=3.000 INCHES

| Diameter | f(mc) | Individual Runs | Average | Calculated |
|------------|-------|-------------------------------------|---------|------------|
| 0.750 inch | 9328 | -15.5 db -15.8 -15.8 -15.6 | -15.7 | -15.54 |
| 0.250 inch | 9330 | -20.9 -20.3 -20.7 -20.9 | -20.7 | -21.07 |

of the target. In some cases, the latter may actually be the limiting and the most crucial factor in the experiment.

CROSS SECTIONS OF CONES

The work which we have done to date has been confined to a family of cones having a semi-apex angle $\alpha = 7.5^\circ$. The cross sections were measured as a function of the altitude L for the two aspects of nose-on incidence (wave incident onto the tip) and base-on incidence. For the nose-on aspect the base plays an important part in the scattering and it was, therefore, interesting to see how perturbations in the base structure affect the backscattering cross section.

Fig. 3 shows how the sequence of targets was built up by adding sections to the base. This procedure was adopted for economy and to insure reproducibility of the region of the tip. The openings in the base sections which are seen in the figure were filled in, of course, by flat-headed screws used to attach successive sections.

The experimental data are shown in Fig. 4.⁹ It is observed that the nose-on aspect yields a cross section which is an oscillating function of L/λ indicating the existence of resonances. The base-on aspect, however, shows a somewhat monotonic dependence on L/λ in the region where the measurements were made.

The several analyses of scattering by cones made by Siegel and Keller, to which we referred earlier, are based on the picture that the total scattering is the result of superposing three contributions: 1) a tip contribution; 2) a slant surface contribution; and 3) edge diffraction and scattering from the base. Concerning the tip and slant surface contributions for the case of nose-on incidence, it is clear that there is no contribution of the form of specularly reflected rays from the cone. Consequently, the actual contribution may be expected to be small. The tip contribution may be evaluated from the rigorous solution for the infinite cone. The slant-surface contribution can be approximated for the case when

⁹ The points at $L/\lambda = 6$ and 7 were measured quickly before the meeting to fill in the curve. They may be in some error as is indicated by possible error brackets.

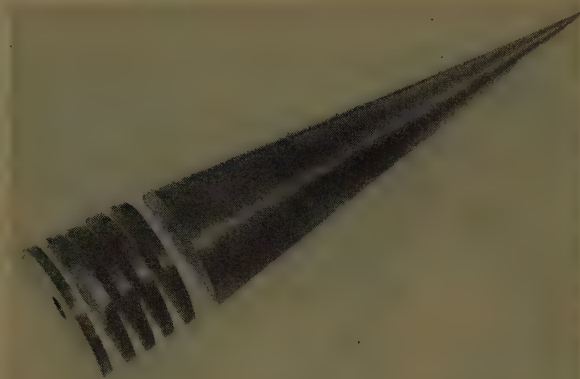


Fig. 3—Finite cone target.

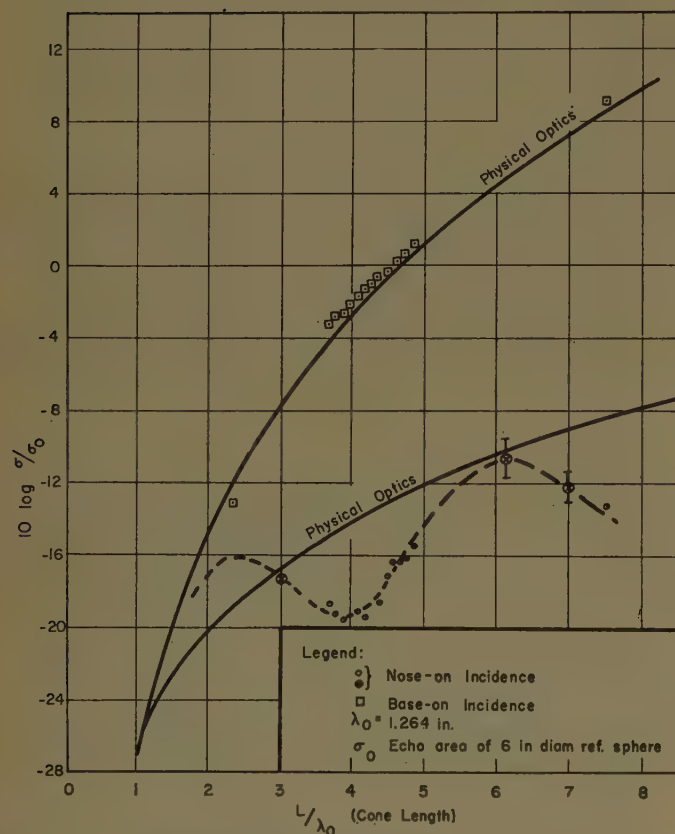


Fig. 4—Backscattering cross section of 15° finite cone as a function of cone length.

$L/\lambda \gg 1$ by finding the radiation field of the so-called geometrical optics current distribution over the surface. The calculations do show that contributions 1) and 2) are small. The slant-surface term shows an oscillating dependence on L/λ which should, however, have disappeared in the region of our datum point ($L/\lambda > 3.5$).

The major contribution is attributable to the base. Siegel's approach is to consider the edge of the base to be a wedge in a local region and to evaluate the scattering from a small arc from the rigorous solution of the problem of diffraction by a wedge. This process of superposition of fields from wedge-like segments yields a cross-

section function which is monotonic in its dependence on L/λ or on a/λ where a is the radius of the base. This cross-section function is represented by the lower of the solid line curves in Fig. 4. The lower portions of the "physical optics" curves below $L/\lambda = 2$ are certainly irrelevant since the physical basis for the analysis is completely nonapplicable there. Keller's approach is to represent the scattering process in terms of families of diffracted rays. His first-order treatment is just like Siegel's in that he obtains the ray system from the edge from the solution of the wedge problem. Superposition of directly scattered rays yields simply the monotonic cross-section function. However, the edge of the base is related to a circular loop and the secondary terms of the interactions between current elements along the edge, which in a general geometrical figure would yield only second-order terms, can add up by virtue of the symmetry to yield first order contributions. The oscillatory behavior of the cross-section function in the region $L/\lambda = 5$ may thus be ascribed to loop-like resonances in the ring of current distribution around the edge. The correspondence between the edge and a loop was exploited by Siegel in discussing the resonance region. In Keller's treatment, the resonances arise as the result of superposition of second-order rays which are generated at one point at the edge of the base and reach the field point by diffraction around the diametrically opposite point. The successive ray development can of course be used also to find the equivalent ring current around the edge. It is interesting to note that two maxima in Fig. 4 correspond to a difference of 3λ between the base circumferences corresponding to those points.

In the case when the primary wave is incident normally onto the base, we have a situation in which there is a dominant body of specularly reflected rays. The cross-section function is, therefore, of the same order as that of a disk. The cross section can be obtained in the form of a high-frequency approximation by using the geometrical optics current distribution over the base. The result is the solid line shown again in Fig. 4 and it is seen to be in agreement with the measured results. The geometrical-optics current distribution approximation should be better for the case of the cone than the disk; for in the case of the cone second-order interactions, which may be thought of in terms of rays passing from one point on the edge to the diametrically opposite point by way of the surface in the shadow region, are attenuated over the long path length, whereas in the case of the disk, we have the situation corresponding to the nose-on incidence for the cone. It will be useful to make a careful set of measurements of scattering by disks in the same range of parameters as that of the cones. The disk should show an extended region of resonances just as the nose-on case of the cone does here.

Returning again to the nose-on aspect of the cone we note that as far as forward scattering is concerned the problem is closely related to that of diffraction by a

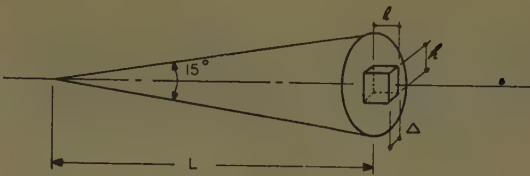
disk. Sometime ago, Ehrlich, Silver, and Held¹⁰ conducted a study of diffraction by circular apertures and the complementary obstacles. They found that in the shadow region of the disk the tangential component of the magnetic field drops rapidly from the edge and rises again to a peak at the center. From the diffracted ray point of view this is simply the effect of in-phase superposition of rays traveling along the surface from the edge. Whatever model one uses, it is clear that here we have a critical region in the current distribution. A change in structure perturbs the distribution, or, if one prefers the ray model again, it scatters the diffracted rays which are again diffracted by the edge, and the scattered field is modified to a proportionately large extent in regions where it is small to begin with. This effect is demonstrated by the results shown in Fig. 5 of a series of measurements of nose-on aspect cross sections as a function of a perturbing structure on the base. The change in cross section is seen to be considerable for some dimensions of the perturbing structure.

PROLATE SPHEROIDS

The family of prolate spheroids which we investigated is shown in Fig. 6. All have an axial ratio (major-to-minor axis) of 10:1. The measurements were made in this case for nose-on incidence (primary wave directed along the major axis) and for broad-side incidence. The spheroids have no singularities, and in both the nose-on and broad-side aspects, we have the condition that a family of specularly reflected rays contribute to the backscattering. The shadow boundary is a region of smooth transition over the surface in contrast with the sharp edge in the case of the cones.

Our results for the nose-on aspect are shown in Table II together with the geometrical-optics values and the values obtained by Siegel and his group¹¹ from the series solution for scattering developed by Schultz.¹² The difficulties associated with the spheroidal functions are well known and beyond the Rayleigh region the problems mount at a rapid rate. Thus the large discrepancy of 22 db between our measured cross section and Siegel's computed cross section for the case $a=2.111$ inches may well be due to computational errors. The agreement between calculated and measured values of the last two in Table II lends confidence to both sources of the cross sections. It must be realized that we are working near the limit of our system where we cannot eliminate the background with the same order of accuracy as in the case of the cones.

¹⁰ M. J. Ehrlich, S. Silver, and C. Held, "Studies of the diffraction of electromagnetic waves by circular apertures and complementary obstacles: the near-zone field," *J. Appl. Phys.*, vol. 26, pp. 336-345; March, 1955.
¹¹ K. M. Siegel, *et al.*, "The theoretical and numerical determination of the radar cross section of a prolate spheroid," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-4, pp. 266-275; July, 1956.
¹² F. V. Schultz, "Scattering by a prolate spheroid," *Engrg. Res. Inst., University of Michigan, Ann Arbor, Mich.*, Rept. No. UMM-42; 1950.



| L (in) | ℓ (in) | Δ (in) | Echo Area in db with Respect to 3" Dia Sphere | Notes |
|--------|--------|--------|---|-----------------------------------|
| 4.60 | 0 | 0 | - 13.2 | Reference at the nose of the cone |
| | 3/64 | 11/16 | - 13.5 | |
| | 1/8 | " | - 14.9 | |
| | 3/16 | " | - 14.2 | |
| | 1/4 | " | - 7.2 | |
| | 3/8 | " | - 1.5 | |
| 9.5 | 0 | 0 | - 4.8 | Reference at the back-face |
| | 1/4 | 1 3/8 | - 7.8 | |

f = 9330 Mc

Fig. 5—Nose-on echo measurements of 15° finite cones with metallic perturbation objects in the shadow.

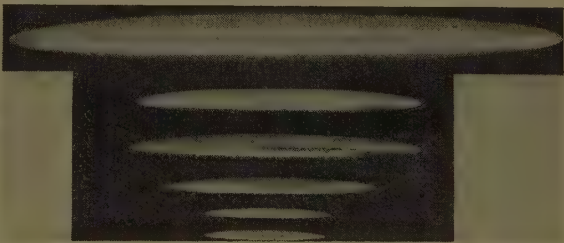


Fig. 6—Prolate spheroid targets.

TABLE II
BACKSCATTERING CROSS SECTIONS OF PROLATE SPHEROIDS OF 10:1 AXIAL RATIO FOR NOSE-ON INCIDENCE WITH RESPECT TO 6-INCH DIAMETER SPHERE (σ_0)

| Semi-Major Axis a (inch) | Geometrical Optics $10 \log \frac{\sigma_{g.o.}}{\sigma_0}$ | Computed by Siegel, <i>et al.</i> | Measured σ $10 \log \frac{\sigma}{\sigma_0}$ |
|--------------------------|--|-----------------------------------|--|
| 6.00 | -33.98 | — | -26.0 |
| 3.00 | -40.0 | — | -48 |
| 2.111 | -43.06 | -65.28 | -43.3 |
| 1.263 | -47.5 | -43.94 | -40.6 |
| 1.184 | -48.06 | -39.21 | -40.9 |

The geometrical optics value given in Table II is obtained from $\pi R_1 R_2$ where R_1 and R_2 are the principal radii of curvature of the surface at the point of normal incidence. The geometrical optics result gives the order of magnitude of the cross section, but the deviations from the values obtained from the series form of the solution and the measured values are significant and are related to the fact that the requisite condition, that the radii are very large compared with wavelength, is not satisfied by the targets we have measured.

Siegel attempted to find a characterization of the resonance region of the scattering function by application of thin wire theory. He used the theory developed by Van Vleck and his co-workers¹³ to determine approximately values of ka ($2\pi a/\lambda$) at which maxima and minima should occur. It is seen from our data that $a=6.00$ inches ($ka=29.8$) is quite likely close to a maximum point in the scattering curve and again that $a=1.184$ ($k=5.9$) is in the neighborhood of a maximum. These agree well with maxima given by thin wire theory.

The cross sections of the spheroids are generally considerably smaller than those for the nose-on aspect of cones. In the present case, we observe that the geometrical-optics contribution from the nose, while not equal to zero, is small and that the contribution of diffracted rays originating at the shadow boundary and traveling over the surface to the far end and around back to the field point is also very small. The diffracted rays are attenuated considerably over the long path over the surface. Since there is no sharp edge to give rise to a more dominant family of rays, the total result is small.

The backscattering cross sections for the case of broadside incidence are shown in Table III. When the spheroid presents its broadside aspect it presents a significant region of specular reflection, considerably more than in the case of nose-on incidence. However, it is seen that even for the largest spheroid the geometrical-optics cross section differs widely from the measured value and that, therefore, we are yet considerably away from the high-frequency limit. In the case $a=6.00$ inches, one of the principal radii of curvature at the point of normal incidence is large compared with the wavelength but the other is comparable to the wavelength. The Gaussian radius of curvature is of the order 1.7λ . Thus diffraction effects are significant. In fact, in the broad-side case, the phase error in the incident illumination is of more consequence than in the nose-on case and it is necessary to measure the cross sections carefully as a function of position in order to determine the actual far zone limit. There is some uncertainty yet in our data in this regard.

It is worth noting in this connection a phenomenon observed by Honda, Silver, and Clapp⁴ in measurements of the cross sections of prolate spheroids as a function of aspect angle. Whereas geometrical optics gives the broad-side aspect the largest cross section it was found that the cross section of the broad-side aspect was slightly less than that of off-set aspects on either side. Honda in his thesis made a detailed analysis on the basis of a true plane wave and showed that the effect is due to the effective phase error introduced over the equivalent aperture for the backscattering by the shape of the surface, that is, the phase distribution of the currents over the spheroid.

¹³ J. H. Van Vleck, F. Bloch and M. Hammermesh, "Theory of radar reflections from wires or thin metallic strips," *J. Appl. Phys.*, vol. 18, pp. 274-294; March, 1947.

TABLE III
BACKSCATTERING CROSS SECTIONS OF PROLATE SPHEROIDS FOR
BROADSIDE INCIDENCE (AXIAL RATIO 10:1) RELATIVE
TO 6-INCH SPHERE

| Semi-Major a (inch) | Geometrical Optics Cross Section $\frac{\sigma_{g.o.}}{\sigma_0}$ $10 \log \frac{\sigma_{g.o.}}{\sigma_0}$ | Measured $10 \log \frac{\sigma}{\sigma_0}$ |
|------------------------|---|---|
| 6.00 | 6.02 | 2.4 |
| 2.993 | 0.00 | -0.5 |
| 2.111 | -3.06 | -4.3 |
| 1.263 | -7.50 | -12.7 |
| 1.184 | -8.06 | -13.5 |

The variation of measured cross section with dimensions shows the same systematic course as the geometrical optics value. There is no evidence of resonance phenomena. We realize, of course, that this may be accidental since our choice of dimensions corresponds to cutting across a possibly oscillating curve at a sequence of points on a monotonic curve. It would be desirable to continue the investigations over a larger range of parameters. We hope to be able to do so in the near future.

OBLATE SPHEROIDS

The family of oblate spheroids which we have measured is shown in Figs. 7 and 8. The idea was to start from a sphere outside the strong resonance region, of diameter 4λ , for instance, and approach the disk through a sequence of oblate spheroids to see whether or not resonance phenomena appear when one dimension gets small. With the two sets of targets shown in the figures, we can measure the cross sections for several different combinations of orientation and polarization.

The results are given in Figs. 9 and 10. In Fig. 9, the circles are cross sections corresponding to the wave incident on the targets shown in Fig. 7. As the minor axis goes to zero, the target reduces to a disk of zero thickness, that is, degenerates into the ground plane surface, and the limiting value of the cross section is zero. The crosses correspond to the family of targets shown in Fig. 8 with the primary wave directed along the major axis. We may term this the "edge-on" aspect corresponding to the limiting case of the disk. Here we observe a relatively large limiting value for the cross section corresponding to the fact that no matter how thin the target may be, a strong current distribution is induced with dominant components normal to the direction of the incident wave. This limiting value is a sensitive function of the diameter of the disk.¹⁴ The solid line in the figure is the geometrical-optics cross section; it is independent of the polarization.

Fig. 10 shows the cross section function for a wave incident on the targets of Fig. 8 in the direction of the minor axis, that is, broad-side incidence. The measured

¹⁴ J. S. Hey, G. S. Stewart, J. J. Pinson, and P. E. Prince, "The scattering of electromagnetic waves by conducting spheres and discs," *Proc. Phys. Soc. (B)*, vol. 69, pp. 1038-1049; October, 1956.

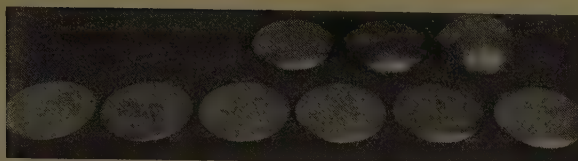


Fig. 7—Oblate spheroid targets.

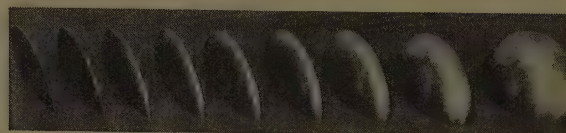


Fig. 8—Oblate spheroid targets.

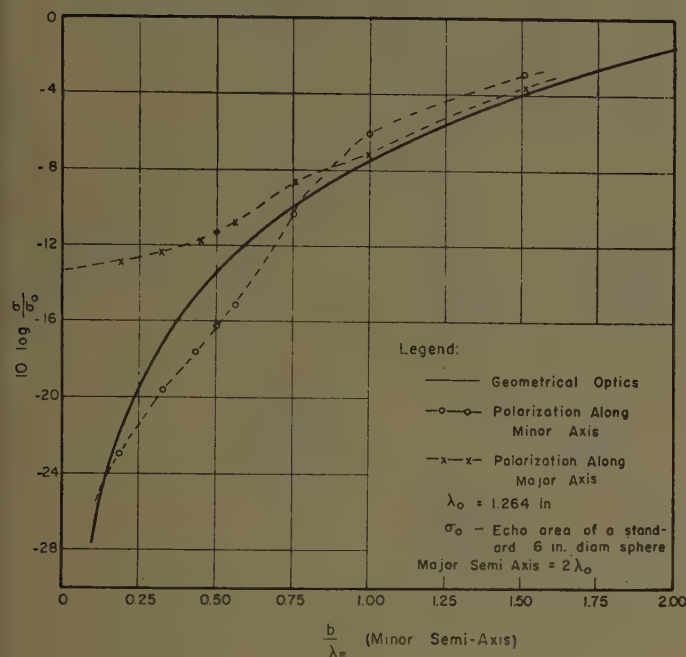


Fig. 9—Backscattering cross sections of oblate spheroids for direction of incidence along major axis.

values show some oscillation about the geometrical-optics curve, but basically follow the curve. The limiting value when $b/\lambda = 2$ must of course correspond to that of the other cases.

The datum points in this series of measurements lie close enough together so that it is hardly likely that the absence of marked oscillation is accidental. With the exception of the series of data for the "edge-on" aspect, the cross sections are sufficiently close to the geometri-

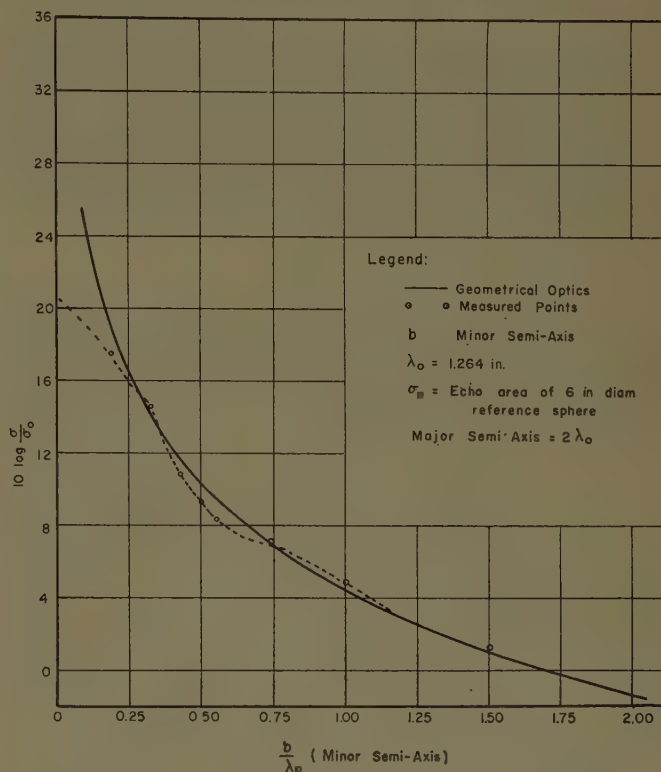


Fig. 10—Backscattering cross section of oblate spheroids for direction of incidence along minor axis.

cal-optics curve that the latter may be considered to be a very good representation of the cross-section function. There are diffraction effects and corresponding deviations from the geometrical-optics functions but the diffraction effect is a small one. How this is accounted for by the diffracted-ray or creeping-wave analysis is not clear, and quantitative studies of those theories applied to targets such as we have investigated are needed very much.

Diffraction by Surfaces of Variable Curvature

W. FRANZ AND K. KLANTE†

Summary—To investigate the influence of variable curvature on the damping of diffracted waves (creeping waves) the integral equation method has been applied to an arbitrary convex cylinder. In addition to the well-known damping factor, depending on the radius of curvature R , a first-order correction term yields an amplitude factor $R^{-1/6}$, while the second-order correction term results in a change in damping depending on the curvature and its first and second derivatives.

I. INTRODUCTION

IT IS known that in the diffraction of waves by a sphere or circular cylinder, the shadowed side of the object is reached by damped waves, called residual waves¹ or creeping waves.²

From the edge of the shadow a series of different creeping waves emerge, which differ in their damping constants. Far from the shadow edge only the first (least damped) creeping wave is left. The damping of the creeping waves per pathlength is proportional to the $\frac{2}{3}$ power of the curvature and to the $-\frac{1}{3}$ power of the wavelength. The theory of the creeping wave may be applied to other than sphere and circular cylinder surfaces by the method of the Maue integral equation.³ If the curvature is not constant along the ray, then it is not possible to have just the one creeping wave of smallest damping corresponding to the uniformly curved surfaces. If, e.g., there were a single discontinuous change of curvature, an incoming creeping wave of the first type would at the discontinuity of curvature create a whole series of creeping waves, belonging to the different value of the curvature. So only after a finite pathlength, would the higher creeping waves be damped out. In effect, the damping would change continuously from the value of the first creeping wave of the initial curvature to that corresponding to the final curvature. If we apply this consideration to a ray of continuously varying curvature, it should be expected that the damping has a certain phase lag as compared to the curvature. To investigate quantitatively the influence of variable curvature on the damping of creeping waves, we start from Maue's integral equation.

† Institute of Applied Physics, University of Münster, Hamburg, Germany.

¹ B. van der Pol and H. Bremmer, "The diffraction of electromagnetic waves from an electrical point source round a finitely conducting sphere," *Phil. Mag.*, vol. 24, pp. 141-176, 825-864; 1937.

² W. Franz and K. Deppermann, "Theorie der Beugung an Zylinder unter Berücksichtigung der Kriechwelle," *Ann. Phys.*, vol. 10, pp. 361-373; 1952.

³ A.-W. Maue, "Zur Formulierung einer allgemeinen Beugungsproblem durch eine Integralgleichung," *Z. Phys.*, vol. 126, pp. 601-608; 1949.

II. DIFFRACTION BY A CONVEX CYLINDER ACCORDING TO MAUE'S INTEGRAL EQUATION

The creeping waves are solutions of the homogeneous Maue equation. For magnetic vector parallel to the cylinder axis the excitation $u(p)$ of the creeping wave at the points P of the surface has to obey

$$u(P) = \frac{i}{2} \int u(Q) \frac{\partial}{\partial n_Q} \mathcal{H}_0^{(1)}(kr) ds_Q. \quad (1)$$

Here s is the pathlength along the ray, r the distance between point of reference P and point of integration Q , and n_Q the outer normal of the surface at Q (see Fig. 1). Now two approximate assumptions are made:

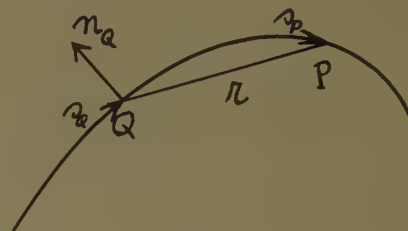


Fig. 1.

the wave number $k = 2\pi/\lambda$ may be large enough, compared to the curvature of the cylinder, to allow for the asymptotic representation of the Hankel function

$$\mathcal{H}_0^{(1)}(kr) \sim \sqrt{\frac{2}{\pi kr}} e^{ikr - i(\pi/4)} \left(1 - \frac{i}{8kr}\right). \quad (2)$$

On the other hand the curvature of the surface shall change slowly enough, to exclude appreciable contributions to the integral (1) from points Q , with a curvature markedly different from that in P . As already known,² the phase of the integral is stationary for points Q , which are reached by the ray just before the point P . Therefore, we may restrict the integration to this region and substitute the integrand by an approximate expression valid in this region. First we split off a factor corresponding to the phase lag along the path

$$u = A(s) \cdot e^{iks}. \quad (3)$$

For the damped amplitude $A(s)$ we write, analogous to the expression known from the circular cylinder,

$$A(s) = \exp \left\{ - \left(\frac{ik}{24} \right)^{1/3} \int_0^s \gamma(\bar{s}) \kappa^{2/3}(\bar{s}) d\bar{s} \right\}. \quad (4)$$

Here κ is the curvature, given as a function of path-length s , and γ is for constant curvature just a numerical constant, for the first creeping wave

$$\gamma(s) = \alpha = 2.33245 \cdot e^{-i\pi/3} \quad \text{for } \kappa \text{ constant.} \quad (5)$$

The purpose of our calculation is to derive the variation of γ as a consequence of variable curvature. By introducing (2)–(4) into (1) we get

$$1 = e^{3\pi i/4} \int_{s_Q} \sqrt{\frac{k}{2\pi r}} \left(1 + \frac{3i}{8kr}\right) \frac{\partial r}{\partial n_Q} \cdot \exp \left\{ ik(s_Q - s_P + r) + \left(\frac{ik}{24}\right)^{1/3} \int_{s_Q}^{s_P} \gamma \kappa^{2/3} ds \right\}. \quad (6)$$

We have confined the integration to the surface in front of P , as the phase becomes stationary only there. Now r , its derivative, and γ are to be expanded into a Taylor series around P . We call

$$\begin{aligned} s_Q - s_P &\equiv s \equiv -s_0; \\ s_0 > 0 &\text{ in stationary phase region;} \end{aligned} \quad (7)$$

and get

$$\begin{aligned} \int_{s_Q}^{s_P} \gamma \kappa^{2/3} ds &= (\gamma \kappa^{2/3})_{P s_0} \\ &- (\gamma \kappa^{2/3})' \cdot \frac{s_0^2}{2} + (\gamma \kappa^{2/3})'' \cdot \frac{s_0^3}{6} \dots \end{aligned} \quad (8)$$

Dashes indicate derivation with respect to s . The Taylor series for r is easily derived by means of the Frenet formulas

$$t' = -\kappa n; \quad n' = \kappa t. \quad (9)$$

Here t is the tangential unit vector, n the unit vector of the outer normal. Obviously t is the derivative of the radius vector

$$t = x'. \quad (10)$$

From (9) and (10) we get by straightforward calculation

$$\begin{aligned} r &= s_0 - \frac{s_0^3}{24} \kappa^2 + \frac{s_0^4}{24} \kappa \kappa' + \frac{s_0^5}{1920} \kappa^4 - \frac{s_0^5}{90} \kappa'^2 \\ &- \frac{s_0^5}{80} \kappa \kappa'' + O(s_0^6). \end{aligned} \quad (11)$$

If we take P as origin of the coordinate system, the normal derivative of r becomes

$$\frac{\partial r}{\partial n_Q} = n_Q \cdot x_Q / r. \quad (12)$$

This yields

$$\frac{\partial r}{\partial n_Q} = \frac{s_0}{2} \kappa - \frac{s_0^2}{3} \kappa' + \frac{s_0^3}{8} \kappa'' - \frac{s_0^3}{48} \kappa^3 + O(s_0^4). \quad (13)$$

We now insert (8), (11), (13) into (6)

$$\begin{aligned} 1 &= e^{3\pi i/4} \sqrt{\frac{k}{8\pi}} \int_0^\infty ds_0 s_0^{1/2} \left(\kappa - \frac{2}{3} \kappa' s_0 \right. \\ &\quad \left. + \frac{\kappa''}{4} s_0^2 - \frac{\kappa^3}{48} s_0^2 \right) \left(1 + \frac{3i}{8k s_0} \right) \\ &\cdot \exp \left\{ -\frac{ik}{24} \kappa^2 s_0^3 \left(1 - \frac{\kappa'}{\kappa} s_0 - \frac{\kappa^2}{80} s_0^2 \right. \right. \\ &\quad \left. \left. + \frac{4}{15} \frac{\kappa'^2}{\kappa^2} s_0^2 + \frac{3}{10} \frac{\kappa''}{\kappa} s_0^2 \right) \right\} \\ &\cdot \exp \left\{ \left(\frac{ik}{24} \right)^{1/3} \left[\gamma \kappa^{2/3} s_0 - \frac{1}{2} (\gamma \kappa^{2/3})' s_0^2 \right. \right. \\ &\quad \left. \left. + \frac{1}{6} (\gamma \kappa^{2/3})'' s_0^3 \right] \right\}. \end{aligned} \quad (14)$$

This is, for $\kappa' = 0$, identical with Franz-Deppermann's integral equation for the circular cylinder. For each of the three factors of the integrand two correction terms containing κ' and κ'' are added. We now introduce the new variable of integration

$$\sigma = s_0 \cdot \left(\frac{ik}{24} \right)^{1/3} \kappa^{2/3} \left(1 - \frac{\kappa'}{\kappa} s_0 - \frac{\kappa^2}{80} s_0^2 \right. \\ \left. + \frac{4}{15} \frac{\kappa'^2}{\kappa^2} s_0^2 + \frac{3}{10} \frac{\kappa''}{\kappa} s_0^2 \right)^{1/3} \quad (15)$$

or (approximately)

$$\begin{aligned} s_0 &= 2 \left(\frac{3}{ik} \right)^{1/3} \kappa^{-2/3} \sigma \cdot \left(1 + \frac{2}{3} \frac{\kappa'}{\kappa^{5/3}} \left(\frac{3}{ik} \right)^{1/3} \sigma \right. \\ &\quad \left. + \left[\frac{\kappa^2}{60} + \frac{44}{45} \frac{\kappa'^2}{\kappa^2} - \frac{2}{5} \frac{\kappa''}{\kappa} \right] \left(\frac{3}{ik} \right)^{2/3} \kappa^{-4/3} \sigma^2 \right); \\ s_0^{1/2} ds_0 &= 2^{3/2} \left(\frac{3}{ik} \right)^{1/2} \kappa^{-1} \sigma^{1/2} d\sigma \left(1 + \frac{5}{3} \frac{\kappa'}{\kappa^{5/3}} \left(\frac{3}{ik} \right)^{1/3} \sigma \right. \\ &\quad \left. + \left[\frac{7}{120} \kappa^2 + \frac{343}{90} \frac{\kappa'^2}{\kappa^2} - \frac{7}{5} \frac{\kappa''}{\kappa} \right] \left(\frac{3}{ik} \right)^{2/3} \kappa^{-4/3} \sigma^2 \right), \end{aligned} \quad (16)$$

and have

$$\begin{aligned} 1 &= i \sqrt{\frac{3}{\pi}} \int_0^\infty d\sigma \cdot \sigma^{1/2} \left(1 + \frac{1}{3} \left(\frac{3}{ik} \right)^{1/3} \frac{\kappa'}{\kappa^{5/3}} \sigma \right. \\ &\quad \left. + \left[-\frac{\kappa^2}{40} + \frac{7}{10} \frac{\kappa'^2}{\kappa^2} - \frac{2}{5} \frac{\kappa''}{\kappa} \right] \left(\frac{3}{ik} \right)^{2/3} \kappa^{-4/3} \sigma^2 \right) \\ &\cdot \left(1 - \frac{1}{16} \left(\frac{3\kappa}{ik} \right)^{2/3} \right) \cdot \exp \left\{ \gamma \sigma - \sigma^3 - \gamma' \left(\frac{3}{ik} \right)^{1/3} \kappa^{-2/3} \sigma^2 \right. \\ &\quad \left. + \gamma \left[\frac{\kappa^2}{60} - \frac{8}{135} \frac{\kappa'^2}{\kappa^2} + \frac{2}{45} \frac{\kappa''}{\kappa} \right] \left(\frac{3}{ik} \right)^{2/3} \kappa^{-4/3} \sigma^3 \right\}. \end{aligned} \quad (17)$$

III. CYLINDER OF SLOWLY VARYING CURVATURE

We now consider a cylinder, which is of slowly varying curvature, allowing for a development of the damping with respect to $(3/ik)^{1/3}$

$$\gamma(s) = \alpha_0 + \alpha_1(s) \left(\frac{3}{ik}\right)^{1/3} + \alpha_2(s) \left(\frac{3}{ik}\right)^{2/3} \dots; \quad (18)$$

α_0 being the pure number of (5). We now develop the integrand of (17) up to second-order terms in $(3/ik)^{1/3}$. If we first take only first-order terms, we have

$$1 = i \sqrt{\frac{3}{\pi}} \int_0^\infty d\sigma \cdot \sigma^{1/2} \cdot e^{\alpha_0 \sigma - \sigma^3} \left(1 + \left(\frac{3}{ik}\right)^{1/3} \left[\frac{1}{3} \frac{\kappa'}{\kappa^{5/3}} + \alpha_1 \right] \sigma \right). \quad (19)$$

As this equation is satisfied, if we take only the 1 of the bracket, it follows that

$$\alpha_1 = -\frac{1}{3} \frac{\kappa'}{\kappa^{5/3}}. \quad (20)$$

Inserting this in (17), the second-order correction α_2 has to be determined from

$$\alpha_2 f''(\alpha_0) = \frac{\kappa^{2/3}}{16} f(\alpha_0) + \left[\frac{\kappa^2}{40} - \frac{4}{45} \frac{\kappa'^2}{\kappa^2} + \frac{1}{15} \frac{\kappa''}{\kappa} \right] \cdot \left(f'''(\alpha_0) - \frac{2}{3} \alpha_0 f''''(\alpha_0) \right). \quad (21)$$

Here $f(\alpha)$ is⁴

$$f(\alpha) \equiv A \left(\frac{e^{i(\pi/3)}}{4^{1/3}} \alpha \right) A \left(\frac{e^{-i(\pi/3)}}{4^{1/3}} \alpha \right) = 4^{-5/6} \sqrt{\frac{\pi}{3}} \int_0^\infty d\sigma \cdot \sigma^{-1/2} \cdot e^{\alpha \sigma - \sigma^3}. \quad (22)$$

A is the Airy integral

$$A(q) \equiv \frac{1}{2} \int_{-\infty}^{+\infty} d\tau \cdot e^{i(q\tau - \tau^3)}. \quad (23)$$

α_0 is a zero of the derivative of the first Airy integral

$$A' \left(\frac{e^{i(\pi/3)}}{4^{1/3}} \alpha_0 \right) = 0. \quad (24)$$

The two Airy integrals of (24) are solutions of the same differential equation.

$$A''(q) + \frac{q}{3} A(q) = 0; \quad (25)$$

where it does not matter which of the arguments is used for q . The Wronskian of the two Airy integrals is the following:

$$e^{-2\pi i/3} A \left(\frac{e^{i(\pi/3)}}{4^{1/3}} \alpha \right) A' \left(\frac{e^{-i(\pi/3)}}{4^{1/3}} \alpha \right) - A' \left(\frac{e^{i(\pi/3)}}{4^{1/3}} \alpha \right) A \left(\frac{e^{-i(\pi/3)}}{4^{1/3}} \alpha \right) = -\frac{\pi}{6} e^{i(\pi/6)}. \quad (26)$$

The differential equation for the function $f(\alpha)$ is

$$f''(\alpha) - \frac{\alpha}{3} f'(\alpha) - \frac{s}{6} f(\alpha) = 0. \quad (27)$$

By virtue of this differential equation α_2 according to (21), contains the fraction

$$\frac{f'''(\alpha_0) - \frac{2}{3} \alpha_0 f''''(\alpha_0)}{f''(\alpha_0)} = \frac{1}{6} \frac{f(\alpha_0)}{f''(\alpha_0)} - \frac{2}{9} \alpha_0^2. \quad (28)$$

Eq. (22), together with (25) and (24), yields

$$f''(\alpha_0) = \frac{1}{6} \alpha_0 f(\alpha_0). \quad (29)$$

So finally we have the result

$$\alpha_2 = \frac{3\kappa^{2/3}}{8\alpha_0} + \left(\frac{1}{\alpha_0} - \frac{2}{9} \alpha_0^2 \right) \cdot \left[\frac{\kappa^2}{40} - \frac{4}{45} \frac{\kappa'^2}{\kappa^2} + \frac{1}{15} \frac{\kappa''}{\kappa} \right] \cdot \kappa^{-4/3}. \quad (30)$$

If we now insert (18), (20), and (30) into (4) the result is

$$A(s) = \left[\frac{\kappa(s)}{\kappa(0)} \right]^{1/6} \exp \int_0^s ds \left\{ - \left(\frac{ik}{24} \right)^{1/3} \alpha_0 \kappa^{2/3} - \frac{3}{16} \left(\frac{3}{ik} \right)^{1/3} \kappa^{4/3} - \frac{1}{720} \left(\frac{3}{ik} \right)^{1/3} \left(\frac{1}{\alpha_0} - \frac{2}{9} \alpha_0^2 \right) \cdot \left[9\kappa^2 - 32 \frac{\kappa'^2}{\kappa^2} + 24 \frac{\kappa''}{\kappa} \right] \kappa^{-2/3} \right\}. \quad (31)$$

For the first creeping wave α_0 has the value of (5). In the coefficient of the second-order term we then have the numerical factor

$$\frac{1}{\alpha_0} - \frac{2}{9} \alpha_0^2 = 1.6377 \cdot e^{i(\pi/3)}. \quad (32)$$

The phase lag of the damping as compared to the curvature, which is to be expected according to our introductory remarks, is contained in the factor $\kappa^{1/6}$. This factor brings an increased damping if the curvature is decreasing, and vice versa. This means that the damping corresponds to a higher respectively lower value of curvature, *i.e.*, to a curvature of a point passed by the ray earlier.

Concerning the validity of the expansion applied, (31) shows that even for constant curvature the product kR (R =radius of curvature) must be large, if the correction term of the exponent is to be omitted; but because of the numerical factors and the power $\frac{1}{3}$, this restriction is rather mild.

$$\left(\frac{kR}{3} \right)^{2/3} > \frac{1}{57}. \quad (33)$$

This, of course, does not tell, whether the higher corrections, which we did not investigate, are also negligible in this case. If the curvature varies, in addition to (33) the relative variation of the curvature and its first derivative must be small within an interval of the order R .

⁴ W. Franz, "Über die Greenschen Funktionen des Zylinders und der Kugel," *Z. Naturforsch.*, vol. 9a, p. 705; 1954.

Diffraction of an Electromagnetic Plane Wave by a Funnel-Shaped Screen

WERNER BRAUNBEK†

Summary—The field on the axis behind a funnel-shaped screen with a circular aperture, on which a plane electromagnetic wave is incident along the axis, is computed by means of an approximate method proposed previously by the author for the case of short wavelengths. The deviation from plane-screen diffraction and the dependence on the angle ϕ_0 defining the shape of the funnel is given for both concave-side and convex-side incidence. Simple closed form results are obtained.

I. INTRODUCTION

THE diffraction of a plane wave incident normally on a plane, infinitesimally thin screen with an aperture of linear dimension a , may be computed approximately for large ka , by adding to the Kirchhoff approximation two terms resulting from the boundary values on two small surface strips along the edge of the aperture. The boundary values, which are really unknown, are replaced by Sommerfeld's boundary values in half-plane diffraction, neglecting the curvature of the edge for sufficiently small wavelengths, *i.e.*, for $ka \gg 1$, where $k = 2\pi/\lambda$ as usual.

This method and its application to the diffraction of a plane scalar wave (sound-wave) by a plane screen with a circular aperture has been described by the author [1], [2].

Franz [3] expanded the same method to the diffraction of an electromagnetic plane wave by a plane screen of the same shape and by a plane circular disk. The results are given in the highest order (zero order) of k and form (in contrast to Kirchhoff's approximation) the rigorous solution of the problem in the limit $ka \rightarrow \infty$.

The same method of approximation may be expanded without difficulties to the diffraction of an electromagnetic plane wave by an aperture in a screen which is not plane. This is done in this paper. In this case one gets only the field on the "shadow" side of the screen, and this only on the conditions that no part of the wave reflected by the illuminated surface comes through the aperture and that no part of the geometrical optics radiation strikes the shadow surface of the screen. The diffraction field behind the screen is, in this approximation, influenced only by the shape of the screen in the vicinity of the edge. In Kirchhoff's approximation the diffraction field is not at all influenced by whether the screen surrounding a plane-edged aperture is plane or not.

Although practical computing is possible only in cases with simple geometry, the importance of the method is based on the fact that until now apertures in nonplane screens were scarcely taken into consideration.

The example shown here is that of computing the electric and the magnetic field on the axis behind a perfectly conducting, infinitely thin, funnel-shaped screen with a circular aperture on which a plane electromagnetic wave is incident along the axis. The conic form of the screen is defined by the angle ϕ_0 (see Figs. 1 and 2).

$\phi_0 > 0$ means incidence of the wave from the concave side, and $\phi_0 < 0$ means incidence from the convex side of the screen. With $\phi_0 = 0$ the problem reduces to that of a plane screen. As for $\phi_0 > \pi/4$ because geometric optics reflected rays would pass through the aperture, ϕ_0 must be restricted to

$$-\frac{\pi}{2} < \phi_0 < +\frac{\pi}{4}.$$

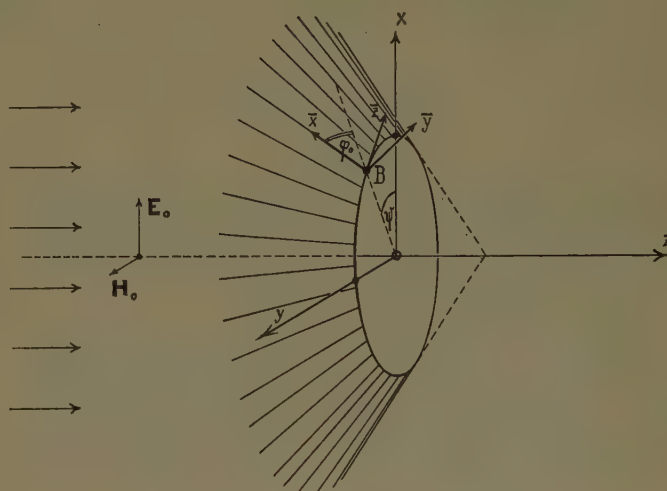


Fig. 1.

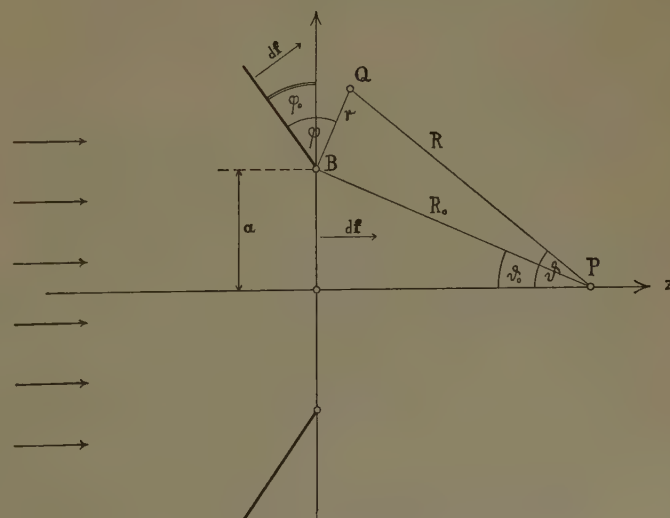


Fig. 2.

† University of Tübingen, Germany.

We assume a plane wave incident along the z axis:

$$E_x = H_y = A e^{ikz} \text{ (time factor } e^{-i\omega t}). \quad (1)$$

All terms and signs are shown in Figs. 1 and 2. In particular, we must express the Sommerfeld half-plane diffraction wave first in the coordinates $\bar{x}\bar{y}\bar{z}$ (and r, ϕ in the $\bar{x}\bar{y}$ plane) individually for each edge-point (defined by the angle ψ), and then transform these coordinates into the common coordinates xyz by means of the transformation formulas for the components of an arbitrary vector \mathbf{V} .

$$\begin{aligned} V_x &= V_{\bar{x}} \cos \phi_0 \cos \psi + V_{\bar{y}} \sin \phi_0 \cos \psi + V_{\bar{z}} \sin \psi, \\ V_y &= V_{\bar{x}} \cos \phi_0 \sin \psi + V_{\bar{y}} \sin \phi_0 \sin \psi - V_{\bar{z}} \cos \psi, \\ V_z &= -V_{\bar{x}} \sin \phi_0 + V_{\bar{y}} \cos \phi_0. \end{aligned} \quad (2)$$

The total field behind the aperture is given by [3], [4]

$$\mathbf{E} = \text{curl} \int (d\mathbf{f} \times \mathbf{E})G + \frac{i}{k} \text{curl} \text{curl} \int (d\mathbf{f} \times \mathbf{H})G \quad (3a)$$

$$\mathbf{H} = \text{curl} \int (d\mathbf{f} \times \mathbf{H})G - \frac{i}{k} \text{curl} \text{curl} \int (d\mathbf{f} \times \mathbf{E})G \quad (3b)$$

with

$$G = \frac{1}{4\pi R} e^{ikR},$$

where R denotes the distance of the field point from the surface element $d\mathbf{f}$ and where the integrals are to be computed over the shadow side of the screen and over the plane surface of the aperture.

From (3) we obtain the Kirchhoff approximation, if we assume in the integrands on the back of the screen $\mathbf{E}=\mathbf{H}=0$, and in the aperture (xy plane) the unchanged values of the incident wave, $E_x=H_y=A$. By this means, one obtains for a point P on the axis in the highest order of k (because of $ka \gg 1$):

$$\begin{aligned} \text{curl}_x \int (d\mathbf{f} \times \mathbf{E})G &= \frac{A}{2} (e^{ikz} - \cos \theta_0 e^{ikR_0}) \\ \frac{i}{k} \text{curl}_x \text{curl} \int (d\mathbf{f} \times \mathbf{H})G &= \frac{A}{2} (e^{ikz} - \frac{1}{2}(1 + \cos^2 \theta_0) e^{ikR_0}) \\ \text{curl}_y \int (d\mathbf{f} \times \mathbf{H})G &= \frac{A}{2} (e^{ikz} - \cos \theta_0 e^{ikR_0}) \\ -\frac{i}{k} \text{curl}_y \text{curl} \int (d\mathbf{f} \times \mathbf{E})G &= \frac{A}{2} (e^{ikz} - \frac{1}{2}(1 + \cos^2 \theta_0) e^{ikR_0}) \end{aligned}$$

and finally

$$\begin{aligned} (E_x)_{\text{Kirchhoff}} &= (H_y)_{\text{Kirchhoff}} \\ &= A(e^{ikz} - \frac{1}{4}(1 + \cos \theta_0)^2 e^{ikR_0}). \end{aligned} \quad (4)$$

In this expression the two parts of (3a) and (3b) originating from the two integrals are not equal as they should be in the case of a plane screen with correct boundary values.

II. THE IMPROVEMENT OF KIRCHHOFF'S APPROXIMATION

We improve Kirchhoff's approximation, which does not contain any influence of the shape of the screen (whether plane or not plane), by adding to the integrands of (3) those terms computed by Sommerfeld's half-plane diffraction which exceed the Kirchhoff boundary values. That is, on the back of the screen we use the full value of Sommerfeld; in the aperture, its surplus over A . These corrections are noticeably different from zero only in two small strips along the edge of the aperture (the breadth of each strip is of the order $1/k$), but they cause, in spite of this, an essential change of the result, even in the highest order (zero order) of k .

With these corrections in the case of a plane screen the two parts of (3a) and of (3b) become equal. It is therefore, in that case, sufficient to compute one part (the simpler one with only one curl-operation) and to double this. In our case, with the nonplanar screen, this simplification does not occur and we have to compute both terms. Because of the curl-curl operation in the second term, it is necessary for obtaining \mathbf{E} and \mathbf{H} on the axis, to compute the integrals not only on the axis, but also in its immediate vicinity up to the second order of x and y .

In the local coordinates $\bar{x}\bar{y}\bar{z}$, the Sommerfeld field components are (the incident wave E component parallel to the edge, the \bar{z} component, is $A \sin \psi$; the parallel H component is $-A \cos \psi$):

$$\begin{aligned} E_{\bar{x}} &= A \cos \psi \left\{ \cos \phi_0 [e^{i\gamma} F(T) - e^{i\gamma'} F(T')] \right. \\ &\quad \left. - 2i \cos \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \sin \frac{\phi}{2} (2kr)^{-1/2} e^{ikr} \right\}, \\ E_{\bar{y}} &= A \cos \psi \left\{ \sin \phi_0 [e^{i\gamma} F(T) + e^{i\gamma'} F(T')] \right. \\ &\quad \left. + 2i \cos \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \cos \frac{\phi}{2} (2kr)^{-1/2} e^{ikr} \right\}, \\ E_{\bar{z}} &= A \sin \psi \{ e^{i\gamma} F(T) - e^{i\gamma'} F(T') \}, \\ H_{\bar{x}} &= A \sin \psi \left\{ \cos \phi_0 [e^{i\gamma} F(T) + e^{i\gamma'} F(T')] \right. \\ &\quad \left. - 2i \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \cos \frac{\phi}{2} (2kr)^{-1/2} e^{ikr} \right\}, \\ H_{\bar{y}} &= A \sin \psi \left\{ \sin \phi_0 [e^{i\gamma} F(T) - e^{i\gamma'} F(T')] \right. \\ &\quad \left. - 2i \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \sin \frac{\phi}{2} (2kr)^{-1/2} e^{ikr} \right\}, \\ H_{\bar{z}} &= -A \cos \psi \{ e^{i\gamma} F(T) + e^{i\gamma'} F(T') \}, \end{aligned}$$

where

$$F(T) = \frac{1}{\sqrt{\pi}} e^{-i(\pi/4)} \int_{-\infty}^T e^{i\tau^2} d\tau; \quad F(+\infty) = 1,$$

$$\gamma = kr \sin(\phi - \phi_0),$$

$$\gamma' = -kr \sin(\phi + \phi_0),$$

$$T = (2kr)^{1/2} \sin \frac{\phi - \phi_0 - \pi/2}{2},$$

$$T' = -(2kr)^{1/2} \sin \frac{\phi + \phi_0 + \pi/2}{2},$$

$$0 \leq \phi \leq 2\pi.$$

As it must be, on the back of the screen ($\phi=0$) we have $E_x=E_z=0$. Moreover we find $H_y=0$ there. We now transform Sommerfeld's field into the coordinates xyz by means of the transformation (2), confine it to the cases $\phi=0$ (back of the screen; index α) and $\phi=\phi_0+\pi$ (aperture plane, see Fig. 2; index β), subtract in the second case Kirchhoff's boundary values (that is, A for E_x and for H_y) and so, obtain the following for the surplus of Sommerfeld's boundary values over Kirchhoff's.

1) On the back of the screen:

$$E_x = A \cos^2 \psi \sin \phi_0 \left\{ 2 \sin \phi_0 \Phi_1(kr) + \sqrt{2} \cos \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \Psi(kr) \right\},$$

$$E_y = A \cos \psi \sin \psi \sin \phi_0 \{ \text{the same bracket} \},$$

$$E_z = A \cos \psi \cos \phi_0 \{ \text{the same bracket} \},$$

$$H_x = -A \cos \psi \sin \psi \left\{ 2 \sin^2 \phi_0 \Phi_1(kr) + \sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \cos \phi_0 \Psi(kr) \right\},$$

$$H_y = A \left\{ 2(\cos^2 \psi + \sin^2 \psi \cos^2 \phi_0) \Phi_1(kr) + \sqrt{2} \sin^2 \psi \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \cos \phi_0 \Psi(kr) \right\},$$

$$H_z = A \sin \psi \sin \phi_0 \left\{ -2 \cos \phi_0 \Phi_1(kr) + \sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \Psi(kr) \right\}.$$

2) In the aperture plane:

$$E_x = -A \left\{ \Phi_2(kr) + [\cos^2 \psi (\cos^2 \phi_0 - \sin^2 \phi_0) + \sin^2 \psi] \Phi_3(kr) - \sqrt{2} \cos^2 \psi \cos \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \cos \frac{\phi_0}{2} \Psi(kr) \right\},$$

$$E_y = A \cos \psi \sin \psi \left\{ 2 \sin^2 \phi_0 \Phi_3(kr) - \sqrt{2} \cos \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \cos \frac{\phi_0}{2} \Psi(kr) \right\},$$

$$E_z = A \cos \psi \left\{ 2 \cos \phi_0 \sin \phi_0 \Phi_3(kr) + \sqrt{2} \cos \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \sin \frac{\phi_0}{2} \Psi(kr) \right\},$$

$$H_x = -A \cos \psi \sin \psi \left\{ 2 \sin^2 \phi_0 \Phi_3(kr) + \sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \sin \frac{\phi_0}{2} \Psi(kr) \right\},$$

$$H_y = A \left\{ -\Phi_2(kr) + [\cos^2 \psi + \sin^2 \psi (\cos^2 \phi_0 - \sin^2 \phi_0)] \Phi_3(kr) - \sqrt{2} \sin^2 \psi \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \sin \frac{\phi_0}{2} \Psi(kr) \right\},$$

$$H_z = -A \sin \psi \left\{ 2 \cos \phi_0 \sin \phi_0 \Phi_3(kr) + \sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \cos \frac{\phi_0}{2} \Psi(kr) \right\}. \quad (5b)$$

In these formulas we used the abbreviations

$$\begin{aligned} \Phi_1(kr) &= e^{-ikr \sin \phi_0} F \left[-(2kr)^{1/2} \sin \left(\frac{\phi_0}{2} + \frac{\pi}{4} \right) \right], \\ \Phi_2(kr) &= F[-(kr)^{1/2}], \\ \Phi_3(kr) &= e^{ikr \sin 2\phi_0} F \left[-(2kr)^{1/2} \sin \left(\phi_0 + \frac{3\pi}{4} \right) \right], \\ \Psi(kr) &= (\pi kr)^{-1/2} e^{i(\pi/4)} e^{ikr}. \end{aligned} \quad (6)$$

In order to substitute (5) into (3), we must compute the vector products in the integrals. As on the screen

$$df_x = df \cos \psi \sin \phi_0$$

$$df_y = df \sin \psi \sin \phi_0$$

$$df_z = df \cos \phi_0,$$

(5a) and in the aperture $df_x=df_y=0$; $df_z=df$, we find the following for these vector products.

1) On the back of the screen:

$$(df \times E) = 0 \quad (E/df),$$

$$(df \times H)_x = dfA \left\{ -2 \cos \phi_0 \Phi_1(kr) + \sqrt{2} \sin^2 \psi \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \Psi(kr) \right\},$$

$$(df \times H)_y = -dfA \sqrt{2} \cos \psi \sin \psi \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \Psi(kr),$$

$$(df \times H)_z = 2dfA \cos \psi \sin \phi_0 \Phi_1(kr); \quad (7a)$$

2) in the aperture plane

$$(df \times E)_x = -dfE_y,$$

$$(df \times E)_y = dfE_x,$$

$$(df \times H)_x = -dfH_y,$$

$$(df \times H)_y = dfH_x,$$

$$z \text{ components zero.} \quad (7b)$$

The magnitude df of the surface element may be approximated by $adr d\psi$ in the two surface strips on which the values (7) differ noticeably from zero.

If we define a field point P' in the close vicinity of the axis by

$$\left. \begin{aligned} x &= p \cos \chi \\ y &= p \sin \chi \end{aligned} \right\} p \text{ infinitesimally small}$$

z arbitrary;

we find the following, if we proceed up to the first order in r , and up to the second order in p , for the distance R of a surface element from P' .

1) For surface elements on the screen:

$$\begin{aligned} R_\alpha &\approx R_0 + r \sin(\theta_0 + \phi_0) - p \sin \theta_0 \cos(\psi - \chi) \\ &+ \frac{p^2}{2R_0} \cos^2 \theta_0 \cos^2(\psi - \chi); \end{aligned}$$

2) for surface elements of the aperture plane:

$$\begin{aligned} R_\beta &\approx R_0 - r \sin \theta_0 - p \sin \theta_0 \cos(\psi - \chi) \\ &+ \frac{p^2}{2R_0} \cos^2 \theta_0 \cos^2(\psi - \chi). \end{aligned}$$

In the expression $G = e^{ikR}/4\pi R$ the surplus of R over R_0 is, in the highest order of k , important only in the exponent; $1/R$ may be substituted by $1/R_0$. So it follows in the highest order of k , if we expand the exponential function with the infinitesimally small p terms into a power series,

$$\begin{aligned} G_\alpha &= \frac{1}{4\pi R_0} e^{ik[R_0 + r \sin(\theta_0 + \phi_0)]} \\ &\times \left\{ 1 - ikp \sin \theta_0 \cos(\psi - \chi) \right. \\ &\quad \left. - \frac{1}{2} k^2 p^2 \sin^2 \theta_0 \cos^2(\psi - \chi) \right\} \\ G_\beta &= \frac{1}{4\pi R_0} e^{ik(R_0 - r \sin \theta_0)} \left\{ \text{the same bracket} \right\}. \quad (8) \end{aligned}$$

From (5), (7), and (8) we see that the ψ dependence in the single terms of (3) may have no other form than

$$1, \quad \cos \psi, \quad \cos^2 \psi, \quad \sin^2 \psi, \quad \cos \psi \sin \psi$$

or products of one of these expressions with $p \cos(\psi - \chi)$ or with $p^2 \cos^2(\psi - \chi)$. The integrals determined by using these terms with ψ from 0 to 2π are:

$$\begin{aligned} \int_0^{2\pi} d\psi &= 2\pi, \\ \int_0^{2\pi} \cos^2 \psi d\psi &= \int_0^{2\pi} \sin^2 \psi d\psi = \pi, \\ p \int_0^{2\pi} \cos \psi \cos(\psi - \chi) d\psi &= \pi x, \\ p^2 \int_0^{2\pi} \cos^2(\psi - \chi) d\psi &= \pi(x^2 + y^2), \\ p^2 \int_0^{2\pi} \cos^2 \psi \cos^2(\psi - \chi) d\psi &= \frac{\pi}{4} (3x^2 + y^2), \\ p^2 \int_0^{2\pi} \sin^2 \psi \cos^2(\psi - \chi) d\psi &= \frac{\pi}{4} (x^2 + 3y^2), \\ p^2 \int_0^{2\pi} \cos \psi \sin \psi \cos^2(\psi - \chi) d\psi &= \frac{\pi}{2} xy. \quad (9) \end{aligned}$$

All other integrals over ψ vanish.

So the integrals occurring in (3) take the following form.

1) On the back of the screen (the E integrals vanish here):

$$\begin{aligned} \int (df \times H)_x G &= \alpha'_0 + \alpha'_{xx} x^2 + \alpha'_{yy} y^2, \\ \int (df \times H)_y G &= \alpha'_{xy} xy, \\ \int (df \times H)_z G &= \alpha'_x x; \end{aligned} \quad (10a)$$

2) in the aperture plane:

$$\begin{aligned} \int (df \times E)_x G &= \beta_{xy} xy, \\ \int (df \times E)_y G &= \beta_0 + \beta_{xx} x^2 + \beta_{yy} y^2, \\ \int (df \times H)_x G &= \beta'_0 + \beta'_{xx} x^2 + \beta'_{yy} y^2, \\ \int (df \times H)_y G &= \beta'_{xy} xy. \end{aligned} \quad (10b)$$

The quantities α', β, β' are integrals over r and depend only on ϕ_0 and on z . They will be given explicitly later on.

We now apply the operations curl and curl-curl to (10) and we make $x=y=0$ after these operations to get the values on the axis. The results are as follows.

1) The contributions of the screen:

$$\text{curl}_y \int (df \times H)G = \frac{\partial \alpha_0'}{\partial z} - \alpha_x'$$

$$\begin{aligned} \text{curl}_x \text{curl} \int (df \times H)G \\ = -\frac{\partial^2 \alpha_0'}{\partial z^2} + \frac{\partial \alpha_x'}{\partial z} + \alpha_{xy}' - 2\alpha_{yy}'; \quad (11a) \end{aligned}$$

2) the contributions of the aperture:

$$\text{curl}_x \int (df \times E)G = -\frac{\partial \beta_0}{\partial z},$$

$$\text{curl}_y \int (df \times H)G = \frac{\partial \beta_0'}{\partial z},$$

$$\text{curl}_y \text{curl} \int (df \times E)G = -\frac{\partial^2 \beta_0}{\partial z^2} - 2\beta_{xx} + \beta_{yy},$$

$$\text{curl}_x \text{curl} \int (df \times H)G = -\frac{\partial^2 \beta_0'}{\partial z^2} + \beta_{xy}' - 2\beta_{yy}'. \quad (11b)$$

All components not listed here vanish on the axis.

The ten values α and β used in (11) are computed from (5) and (7)–(9) as follows:

$$\begin{aligned} \alpha_0' &= \frac{A}{4k} \sin \theta_0 e^{ikR_0} \left\{ -4 \cos \phi_0 \bar{\Phi}_1 \right. \\ &\quad \left. + \sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \bar{\Psi}_\alpha \right\}, \end{aligned}$$

$$\alpha_x' = -\frac{iA}{2} \sin^2 \theta_0 e^{ikR_0} \sin \phi_0 \bar{\Phi}_1,$$

$$\alpha_{xy}' = \frac{kA}{16} e^{ikR_0} \sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \bar{\Psi}_\alpha,$$

$$\alpha_{yy}' = -\frac{kA}{32} \sin^3 \theta_0 e^{ikR_0} \left\{ -8 \cos \phi_0 \bar{\Phi}_1 \right.$$

$$\left. + 3\sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \bar{\Psi}_\alpha \right\},$$

$$\beta_0 = -\frac{A}{4k} \sin \theta_0 e^{ikR_0} \left\{ 2\bar{\Phi}_2 + 2 \cos^2 \phi_0 \bar{\Phi}_3 \right.$$

$$\left. + \sqrt{2} \cos \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \cos \frac{\phi_0}{2} \bar{\Psi}_\beta \right\},$$

$$\beta_{xx} = \frac{kA}{32} \sin^3 \theta_0 e^{ikR_0} \left\{ 4\bar{\Phi}_2 + (4 \cos^2 \phi_0 - 2 \sin^2 \phi_0) \bar{\Phi}_3 \right.$$

$$\left. + 3\sqrt{2} \cos \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \cos \frac{\phi_0}{2} \bar{\Psi}_\beta \right\},$$

$$\beta_{xy} = \frac{kA}{16} \sin^3 \theta_0 e^{ikR_0} \left\{ 2 \sin^2 \theta_0 \bar{\Phi}_3 \right.$$

$$\left. - \sqrt{2} \cos \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \cos \frac{\phi_0}{2} \bar{\Psi}_\beta \right\},$$

$$\begin{aligned} \beta_0' &= -\frac{A}{4k} \sin \theta_0 e^{ikR_0} \left\{ -2\bar{\Phi}_2 + 2 \cos^2 \phi_0 \bar{\Phi}_3 \right. \\ &\quad \left. - \sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \sin \frac{\phi_0}{2} \bar{\Psi}_\beta \right\}, \end{aligned}$$

$$\begin{aligned} \beta_{xy}' &= \frac{kA}{16} \sin^3 \theta_0 e^{ikR_0} \left\{ 2 \sin^2 \phi_0 \bar{\Phi}_3 \right. \\ &\quad \left. + \sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \sin \frac{\phi_0}{2} \bar{\Psi}_\beta \right\}, \end{aligned}$$

$$\begin{aligned} \beta_{yy}' &= \frac{kA}{32} \sin^3 \theta_0 e^{ikR_0} \left\{ -4\bar{\Phi}_2 + (4 \cos^2 \phi_0 - 2 \sin^2 \phi_0) \bar{\Phi}_3 \right. \\ &\quad \left. - 3\sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right) \sin \frac{\phi_0}{2} \bar{\Psi}_\beta \right\}. \quad (12) \end{aligned}$$

In these formulas we used the notations (with $kr = \zeta$):

$$\bar{\Phi}_1 = \int_0^\infty \Phi_1 e^{i\zeta \sin(\theta_0 + \phi_0)} d\zeta,$$

$$\bar{\Phi}_2 = \int_0^\infty \Phi_2 e^{-i\zeta \sin \theta_0} d\zeta,$$

$$\bar{\Psi}_\alpha = \int_0^\infty \Psi_\alpha e^{i\zeta \sin(\theta_0 + \phi_0)} d\zeta,$$

$$\bar{\Psi}_\beta = \int_0^\infty \Psi_\beta e^{-i\zeta \sin \theta_0} d\zeta.$$

These integrals should be taken over r beginning with $r=0$, only over a small interval corresponding to the small size of the surface strips. As the integrands approach zero rather quickly outside of this interval, it is permissible to integrate from 0 to ∞ . So the integrals may be evaluated in closed form and one gets [2]

$$\bar{\Phi}_1 = \frac{i}{2[\sin(\theta_0 + \phi_0) - \sin \phi_0]} \left[1 - \frac{\sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{\pi}{4} \right)}{\sqrt{1 + \sin(\theta_0 + \phi_0)}} \right],$$

$$\bar{\Phi}_2 = -\frac{i}{2 \sin \theta_0} \left[1 - \frac{1}{\sqrt{1 - \sin \theta_0}} \right],$$

$$\bar{\Phi}_3 = \frac{i}{2(\sin 2\phi_0 - \sin \theta_0)} \left[1 - \frac{\sqrt{2} \sin \left(\frac{\phi_0}{2} + \frac{3\pi}{4} \right)}{\sqrt{1 - \sin \theta_0}} \right],$$

$$\bar{\Psi}_\alpha = \frac{i}{\sqrt{1 + \sin(\theta_0 + \phi_0)}},$$

$$\bar{\Psi}_\beta = \frac{i}{\sqrt{1 - \sin \theta_0}}. \quad (13)$$

Now we have only to substitute (13) into (12), (12) into (11), and (11) into (3) in order to find the additional terms of \mathbf{E} and \mathbf{H} on the axis which together with the Kirchhoff terms (4) form the total field. In (11) however, we have to differentiate with respect to z . As all functions to be differentiated have the form $f(\phi_0) e^{ikR_0}$, in the highest order of k , differentiation in the exponent is sufficient.

$$\frac{\partial}{\partial z} [f(\theta_0) e^{ikR_0}] \approx ik \frac{\partial R_0}{\partial z} f(\theta_0) e^{ikR_0} = ik \cos \theta_0 f(\theta_0) e^{ikR_0}.$$

That means that differentiation with respect to z may be replaced by multiplication with $ik \cos \theta_0$. With this done one gets for the total field on the axis behind the screen

$$\begin{aligned} E_x &= A \{ e^{ikz} - e^{ikR_0} F_1(\theta_0, \phi_0) \} \\ H_y &= A \{ e^{ikz} - e^{ikR_0} F_2(\theta_0, \phi_0) \}. \end{aligned} \quad (14)$$

By straightforward computation, the functions F_1 and F_2 are found to be very complicated expressions. By passing to the angle $\theta_0/2$ however, and by elementary but tedious transformations, one succeeds in getting the simple form

$$F_{\frac{1}{2}} = \cos^2 \frac{\theta_0}{2} \pm \sin^2 \frac{\theta_0}{2} \cdot \frac{\cos \frac{\theta_0}{2}}{\cos \left(\frac{\theta_0}{2} + \phi_0 \right)}. \quad (15)$$

The angle ϕ_0 characterizing the funnel shape of the screen occurs in (15) only in the denominator of the second term. For the case of a circular aperture in a plane screen, (14) with (15) reduces, as $\cos \phi_0/2 \pm \sin \phi_0/2 = (1 \pm \sin \phi_0)^{1/2}$, to

$$\begin{aligned} (E_x)_{\text{plane screen}} &= A \{ e^{ikz} - e^{ikR_0} (1 - \frac{1}{2} \sin \theta_0) \sqrt{1 + \sin \theta_0} \} \\ (H_y)_{\text{plane screen}} &= A \{ e^{ikz} - e^{ikR_0} (1 + \frac{1}{2} \sin \theta_0) \sqrt{1 - \sin \theta_0} \}. \end{aligned}$$

This coincides with Franz's solution [3] with the exception that in Franz's solution the transmitted wave e^{ikz} is missing and that, moreover, \mathbf{E} and \mathbf{H} are interchanged (with a change of sign). Both facts result from Franz's problem, the diffraction by a circular disk, being the complementary problem to ours (in the special case of a plane screen) and being connected to ours by Babinet's principle.

The final result (14) with (15) is valid for

$$-\pi/2 < \phi_0 < +4/\pi,$$

that is only up to 45° for cones on which the incident wave strikes from the concave side. For the other direction of incidence the formulas are valid up to 90° (cylinder). In the vicinity of each limit the approximation may be rather poor.

According to the computation followed, the result should be valid for $0 < \theta_0 < \pi/2$, that is, for all points on the axis on the shadow side of the screen. In a former paper [5] however, the author demonstrated why the approximate method used here does not hold in the immediate vicinity of the center of the aperture ($\theta_0 = \pi/2$) and showed that the following restriction applies:

$$ka(1 - \sin \theta_0) \gg 1$$

or

$$\pi/2 - \theta_0 \gg (ka)^{-1/2}.$$

This restriction is less drastic the larger ka is, and it vanishes with $ka \rightarrow \infty$.

For very remote points of the axis, $\phi_0 \rightarrow 0$, the field approaches

$$E_x \rightarrow A(e^{ikz} - e^{ikR_0})$$

$$H_y \rightarrow A(e^{ikz} + e^{ikR_0}), \text{ independent of } \phi_0.$$

The main results are demonstrated in Figs. 3 and

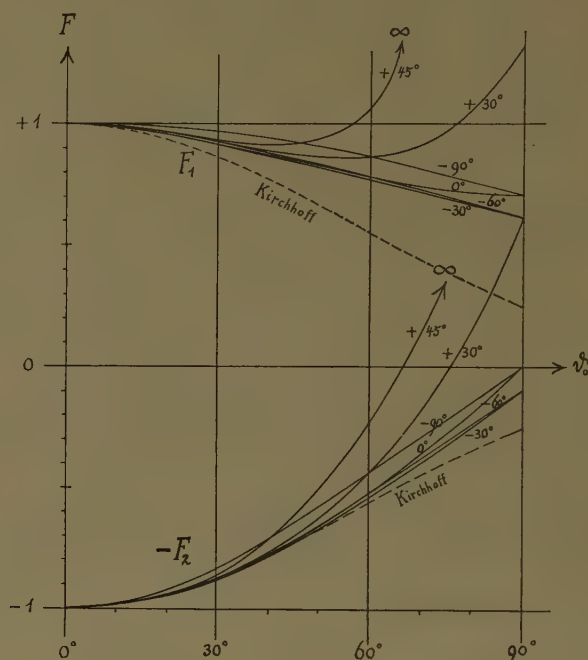


Fig. 3.

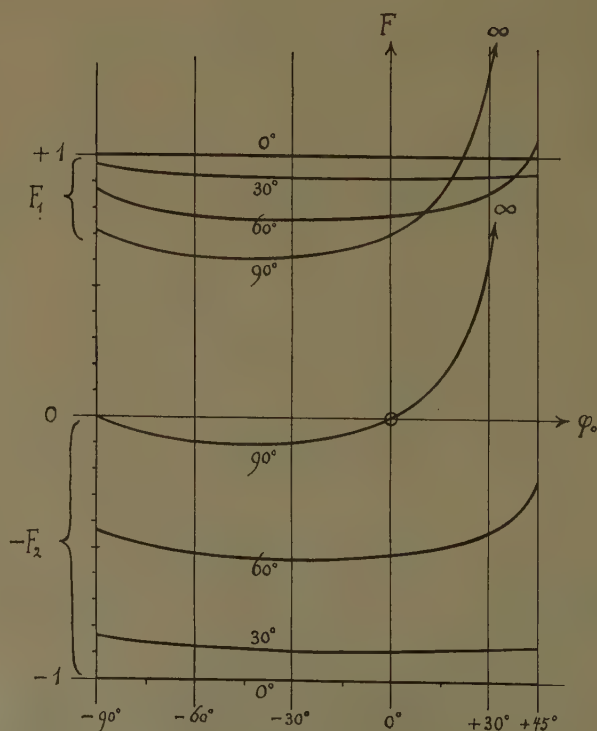


Fig. 4.

4. F_1 and $(-F_2)$, according to (15), are shown in Fig. 3 as functions of θ_0 with parameter ϕ_0 and in Fig. 4 as functions of ϕ_0 with parameter θ_0 . The representation in Fig. 3 is less clear because of the numerous crossings of the closely lying curves. The dotted lines are Kirchhoff's approximation. All curves of Fig. 4 are symmetrical about the points $\phi_0 = -\theta_0/2$, which follows from the form of (15). The curves show clearly that a positive angle ϕ_0 (wave incident from the concave side of the screen) causes a larger deviation from the plane-screen case than a negative ϕ_0 of the same magnitude.

BIBLIOGRAPHY

- [1] W. Braunbek, "Neue Näherungsmethode für die Beugung am ebenen Schirm," *Z. für Phys.*, vol. 127, pp. 381-390; 1950.
- [2] ———, "Zur Beugung an der Kreisscheibe," *Z. für Phys.*, vol. 127, pp. 405-415; 1950.
- [3] W. Franz, "Theorie der Beugung elektromagnetischer Wellen," Springer-Verlag, Berlin, Ger., pp. 103-109; 1957.
- [4] C. J. Bouwkamp, "Diffraction theory," *Reps. Prog. Phys.*, vol. 17, p. 57 (5.6); 1954.
- [5] W. Braunbek, "Zur Beugung an der kreisförmigen Öffnung," *Z. für Phys.*, vol. 138, pp. 80-88; 1954.
- [6] J. B. Keller, "Diffraction by an aperture," *J. Appl. Phys.*, vol. 28, pp. 426-444; 1957.
- [7] J. B. Keller, R. M. Lewis, and B. D. Seckler, "Diffraction by an aperture II," *J. Appl. Phys.*, vol. 28, pp. 570-579; 1957.

The Experimental Determination of the Far-Field Scattering from Simple Shapes*

J. E. KEYS† AND R. I. PRIMICH†

ABSTRACT

THREE model radar ranges have been constructed to measure the scattering cross-section of models. The first of these is a 10- μ sec pulse radar operating at a frequency of 35 kmc. The second, operating at the same frequency, is a CW balanced hybrid T system. The third, at 8.75 kmc, has two modes of operation.

In one it operates as a simple CW balanced hybrid T system and in the other, which is inherently more stable, it uses low frequency modulation to achieve a range resolution of about 7 feet.

As a combined facility, with various sizes of model, we can study the back-scatter cross-section of simple models, for any polarization, over an effective wavelength range of at least 40 to 1. The minimum detectable cross-section is about 45 db below a square wavelength.

Back-scatter measurements have been made on a variety of simple shapes. Results were presented on metal cones and toroids.

* A report covering the essential material of this paper will be published at a later date.

† Defence Res. Telecommun. Est., Defence Research Board, Ottawa, Can.

The Diffraction and Refraction of Plane Pulses*

V. M. PAPADOPOULOS†

Summary—A method for solving problems of plane pulse diffraction by a perfectly conducting wedge is described. The method is extended to give results when a conductive half-plane lies on the surface between two distinct isotropic media of different dielectric properties.

I. INTRODUCTION

BOTH in the theory of acoustics which describes the propagation of infinitesimal disturbances, and in the closely related electromagnetic theory, research into the propagation of aperiodic disturbances has been overshadowed by research into the behavior of disturbances varying sinusoidally with time. It is possible to use a Fourier integral method to investigate aperiodic disturbances within a single homogeneous medium, but Craggs [1], [2] points out that such a method is useless in cases when total reflexion is to be expected at the surface of discontinuity between two homogeneous media.

Friedlander [3] has recently published an excellent monograph in which he discusses the behavior of acoustic pulses. He does not, however, give much attention to the refraction of pulses, in spite of having derived [4] an interesting result in this subject. We already know what combination of reflected and refracted pulses can travel along an infinite plane surface between two media with an incident pulse having a step-function time dependence. Friedlander found the nature of the disturbance in the second medium when the angle of incidence of the pulse in the first medium is large enough for total reflexion to occur. The manner in which such a steadily traveling disturbance can be set up is not known.

The first aim of this research is to learn how a plane pulse in one medium sets up a disturbance in a second medium, even in the case of total reflexion. The second aim is to find a solution to the problem of diffraction in the presence of a refracting surface. There is no known solution for this problem in the case of steady sinusoidal excitation, even in the simplest extension of the classical problem of half-plane diffraction, when we take the obstacle to lie on the surface separating two distinct homogeneous regions. The powerful method of Wiener and Hopf gives no usable result.

Craggs [1], [2] has shown that an assumption of dynamic similarity may be useful in this type of problem. Thus when investigating a scalar function $s(r, \theta, t)$ which satisfies within a single medium the wave equation

$$c^2 \nabla^2 s = \partial^2 s / \partial t^2, \quad (1)$$

where c is the constant velocity of propagation, we may take $\lambda = r/t$ to be a new independent variable. Then (1) becomes

$$\lambda^2 \left(1 - \frac{\lambda^2}{c^2}\right) \frac{\partial^2 s}{\partial \lambda^2} + \lambda \left(1 - \frac{2\lambda^2}{c^2}\right) \frac{\partial s}{\partial \lambda} + \frac{\partial^2 s}{\partial \theta^2} = 0. \quad (2)$$

For $\lambda > c$, (2) is hyperbolic; it may be reduced to the canonical form

$$\frac{\partial^2 s}{\partial \mu^2} - \frac{\partial^2 s}{\partial \theta^2} = 0$$

by the transformation $\lambda = c \sec \mu$. The general solution of this equation is

$$s = f(\mu - \theta) + g(\mu + \theta), \quad (3)$$

where f and g are arbitrary functions, constant on the family of characteristics $\mu - \theta = \text{constant}$, and $\mu + \theta = \text{constant}$, respectively. In the plane with polar coordinates (λ, θ) these characteristics are tangents to the circle $\lambda = c$; it is necessary to regard the parts of a tangent on opposite sides of the point of contact as distinct and of opposite family.

For $\lambda < c$, (2) is elliptic. It reduces to

$$\frac{\partial^2 s}{\partial \nu^2} + \frac{\partial^2 s}{\partial \theta^2} = 0, \quad (4)$$

under the transformation $\lambda = c \operatorname{sech}(-\nu)$. The sign of ν is chosen so that $\nu \rightarrow \ln(\lambda/c)$ as $\lambda \rightarrow 0$. It is convenient to introduce the conjugate harmonic function $\tau(\nu, \theta)$ so that the function $w = s + i\tau$ is analytic in the region of the (ν, θ) plane which corresponds to $\lambda < c$. It follows that $\partial s / \partial \theta = -\partial \tau / \partial \nu$, and $\partial s / \partial \nu = \partial \tau / \partial \theta$.

Eqs. (3) and (4) show the technical start in the solution of a number of physical problems. In electromagnetic theory, by considering problems in cylindrical polar coordinates (r, θ, z) in which the field components are independent of z , $s(r, \theta, t)$ may represent either the field component E_z or B_z . From Maxwell's equations we can write down equations for the remaining field components in both an E and an H polarization. These are

$$\lambda^2 \frac{\partial B_r}{\partial \lambda} = \frac{\partial E_\theta}{\partial \theta}, \quad -\lambda \frac{\partial B_\theta}{\partial \lambda} = \frac{\partial E_r}{\partial \lambda}, \quad (E\text{-mode}); \quad (5)$$

and

$$\lambda^2 \frac{\partial E_r}{\partial \lambda} = -c^2 \frac{\partial B_\theta}{\partial \theta}, \quad \lambda \frac{\partial E_\theta}{\partial \lambda} = c^2 \frac{\partial B_r}{\partial \lambda}, \quad (H\text{-mode}). \quad (6)$$

* This research is sponsored in part by the AF Cambridge Research Center, under Contract AF 19(604)-4561, and in part by the Office of Naval Research and the David W. Taylor Model Basin, under Contract Nonr-562(24).

† Brown University, Providence, R. I.

The problems to be examined all involve infinite prisms with the line $r=0$ for apex. We set up a system of plane pulses traveling towards the apex so that behind the pulse fronts all the appropriate boundary conditions are satisfied. We take the moment when the incident pulses reach the apex to be $t=0$, and we assume, in the absence of a fundamental length in the geometry, that the subsequent disturbance is one in which there is dynamic similarity. We take the prism to be bounded externally by perfectly conducting walls.

First we shall derive, in Section II, the results for pulse diffraction within a homogeneous prism; these results may be compared with those established previously (e.g., [3]).

We shall then extend the method in Section III to find a solution in the case in which a half-plane with its edge at the origin lies on the surface which separates two distinct isotropic media. We set as the restriction on the angle of incidence the condition that the incident pulses shall be contained in one medium only.

II. PULSE DIFFRACTION IN A HOMOGENEOUS MEDIUM BY A HALF-PLANE AND BY A WEDGE

Consider the case of a perfectly conducting half-plane in the diffraction problem where $s=B_z$, and the appropriate boundary condition on the half-plane is that $\partial s/\partial\theta=0$. We set up a plane pulse traveling with the velocity c , and with a step-function time dependence of unit amplitude. There are two cases which need separate examination: 1) with the pulse traveling towards the edge of the half-plane from the unbifurcated region at an arbitrary angle, and 2) with the pulse traveling along the half-plane towards the edge, at oblique incidence. We take the moment when the incident pulse meets the edge to be $t=0$. The distinction between the two cases is needed, since, in order to satisfy the boundary condition on the half-plane for $t<0$ in Case 1, we must introduce the associated reflected pulse. The conditions on our problem at $t=0$ are shown in Fig. 1. In each case we take the incident pulse to make an angle α with the normal to the half-plane with $\alpha\leq\pi/2$. We fix the coordinates by taking the origin at the edge of the half-plane $\theta=0$ or $\theta=2\pi$.

Under the assumption of dynamic similarity, we may examine the nature of the solution for $\lambda>c$ in the (λ, θ) plane. Since for $t\rightarrow 0_+$, $\lambda\rightarrow\infty$, we find that the initial conditions of the problem fix the boundary conditions on (3) for $\lambda\rightarrow\infty$. The use of characteristics theory shows us that for $\lambda>c$ there are a number of uniform regions separated by lines of discontinuity which are depicted in Fig. 2. The results are precisely those of geometrical optics. The lines of discontinuity are pulse fronts. These results determine the value of s on $\lambda=c$.

Thus on $\lambda=c$,

$$\begin{aligned} s &= 2 & \text{for } 0 < \theta < \beta, \\ s &= 1 & \text{for } \beta < \theta < \gamma, \\ s &= 0 & \text{for } \gamma < \theta < 2\pi, \end{aligned}$$

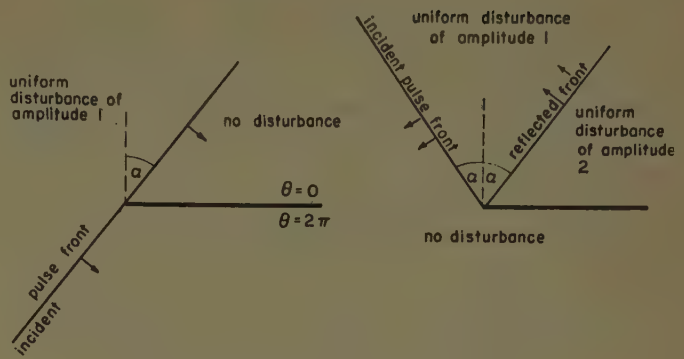


Fig. 1—Initial states in the diffraction of a pulse by a reflecting half-plane.

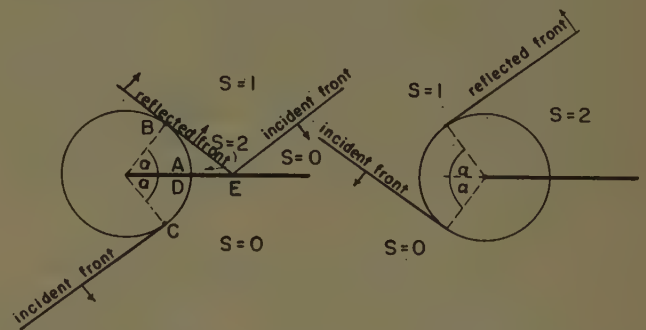


Fig. 2—The solution in the hyperbolic region of the (λ, θ) plane

where $\beta=\alpha$ in Case 1, $\beta=\pi-\alpha$ in Case 2, and $\gamma=2\pi-\beta$.

For the region $\lambda<c$ it follows from (4) that the problem becomes that of finding a function $w(\nu+i\theta)=s+it$ which satisfies the boundary condition $\partial s/\partial\theta=0$ on $\theta=0$, $\theta=2\pi$, and on $\nu=0$ (i.e., $\lambda=c$); and which is analytic within the semi-infinite strip $\nu>0$, $0<\theta<2\pi$. We make the conformal transformation $\zeta=\text{sech } \nu+i\theta/2$, which maps this semi-infinite strip into the upper half of the complex ζ plane. Then the function $\partial\omega/\partial\zeta$, regarded as a function of the complex variable $\zeta=\xi+i\eta$ must satisfy the following conditions:

1) To correspond to the boundary condition on the half-plane, it follows that $\partial\omega/\partial\zeta$ is real on the segment of the real axis $|\xi|<1$.

2) To correspond to the boundary condition that s is piecewise constant on the circle $\lambda=c$, it follows that $\partial\omega/\partial\zeta$ is imaginary on the remainder of the real axis ($|\xi|>1$), except perhaps at points where s is discontinuous.

3) Singularities of $\partial\omega/\partial\zeta$ are to be expected only at the points $\zeta=\pm 1$, $\zeta=0$, as well as at the points $\zeta=\sec \alpha/2$, $\sec (2\pi-\alpha)/2$ in Case 1 and at $\sec (\pi+\alpha)/2$ in Case 2.

4) Since on physical grounds the magnetic field in this polarization is assumed finite, no singularity of $\partial\omega/\partial\zeta$ may be of higher order than a simple pole.

5) The point at infinity in the ζ plane corresponding to the ordinary point $\lambda=c$, $\theta=\pi$ in the (λ, θ) plane must be an ordinary point for ω and $\partial\omega/\partial\zeta$, so that $\omega=0(1/\zeta^{1+\delta})$, with $\delta>0$ as $\zeta\rightarrow\infty$.

6) There is a geometrical singularity at the edge of the half-plane so that we may expect a singularity in $\partial\omega/\partial\zeta$ at the origin $\zeta=0$. It is reasonable however to assume that there is no source of energy at the edge; it follows that $\partial\omega/\partial\zeta=0(\zeta^\delta)$ as $\zeta\rightarrow 0$ where $\delta>-1$ is the smallest permissible value which is compatible with the other conditions of the problem.

The most general function which satisfies these conditions is

$$\pi \frac{\partial\omega}{\partial\zeta} = \frac{i}{(\zeta^2 - 1)^{1/2}} \left[\frac{B'}{\zeta - \sec \beta/2} + \frac{C'}{\zeta - \sec \gamma/2} \right], \quad (7)$$

where B' and C' are real constants which are determined by the discontinuities in s at the points B and C in the (λ, θ) plane. To find the constant B' , we integrate $\partial\omega/\partial\zeta$ round the small semicircle $\zeta = \sec \beta/2 + \exp i\phi$, $0 \leq \phi \leq \pi$, by the method of residues; we find C' in a similar manner. It follows that $B' \cot \beta/2$, $C' \cot \gamma/2$ must equal the discontinuity in s on the circle at the points B and C respectively taken in the direction of increasing θ .

Hence,

$$\pi \frac{\partial\omega}{\partial\zeta} = \frac{-i}{(\zeta^2 - 1)^{1/2}} \left[\frac{\tan \beta/2}{\zeta - \sec \beta/2} + \frac{\tan \gamma/2}{\zeta - \sec \gamma/2} \right]. \quad (8)$$

The integration of this equation is straightforward. We find, after putting $\gamma = 2\pi - \beta$, that

$$\begin{aligned} \pi[\omega(\nu, \theta) - \omega(0, \theta)] \\ = -i \ln \left\{ \frac{\tan \left(\frac{\beta + \theta - i\nu}{4} \right) \tan \left(\frac{\beta - \theta}{4} \right)}{\tan \left(\frac{\beta + \theta}{4} \right) \tan \left(\frac{\beta - \theta + i\nu}{4} \right)} \right\}. \end{aligned}$$

The real part of this equation is

$$\begin{aligned} \pi[s(\nu, \theta) - s(0, \theta)] = -\tan^{-1} \left[\sinh \frac{\nu}{2} \operatorname{cosec} \frac{\beta - \theta}{2} \right] \\ - \tan^{-1} \left[\sinh \frac{\nu}{2} \operatorname{cosec} \frac{\beta + \theta}{2} \right]. \quad (9) \end{aligned}$$

This formula for the diffracted field is exactly that derived and described by Friedlander¹ using a Green's function method.

The results for the E polarization may be found just as easily. For the hyperbolic region in the (λ, θ) plane it is clear that the reflected pulse must annul the incident field instead of reinforcing it. Hence, with reference to Fig. 2, the region with $s=2$ becomes a region with $s=0$, and the discontinuity in s at B is changed accordingly. For the elliptic region, we need only change the condition 1). On the half-plane s must vanish, so that its tangential derivative is zero. It follows that on the seg-

ment of the real ζ axis $|\xi| < 1$, $\partial\omega/\partial\zeta$ is imaginary.

The function satisfying the new set of conditions and with the correct discontinuities is

$$\pi \frac{\partial\omega}{\partial\zeta} = \frac{i}{\zeta} \left[\frac{\sec \beta/2}{\zeta - \sec \beta/2} - \frac{\sec \gamma/2}{\zeta - \sec \gamma/2} \right]. \quad (10)$$

Hence, the final expression for s is

$$\begin{aligned} \pi[s(\nu, \theta) - s(0, \theta)] = \tan^{-1} \left[\sinh \frac{\nu}{2} \operatorname{cosec} \frac{\beta - \theta}{2} \right] \\ - \tan^{-1} \left[\sinh \frac{\nu}{2} \operatorname{cosec} \frac{\beta + \theta}{2} \right]. \quad (11) \end{aligned}$$

The analysis of the problem of pulse diffraction by a perfectly reflecting wedge is carried out in the same way. We exclude, from this brief discussion, any case in which the incident front touches both walls of the wedge, in order to avoid the complication of multiple reflexion before the apex is reached. This leaves us with the general problem for which the wedge angle $\phi\pi$ is greater than $\pi/2$, and the angle $\alpha\pi$ which an incident pulse makes with the normal to one face of the wedge is smaller than $\pi(\phi - \frac{1}{2})$.

In such a situation we depict an initial state in Fig. 3. We use the method of characteristics to determine the solution in the hyperbolic region. This is depicted in Fig. 4, hence we note that on the optic circle $\lambda = c$.

$$s = 4 \quad \text{for} \quad 0 < \theta < \pi(1 - \alpha - \phi),$$

$$s = 3 \quad \text{for} \quad \pi(1 - \alpha - \phi) < \theta < \pi(1 + \alpha - \phi),$$

and

$$s = 2 \quad \text{for} \quad \pi(1 + \alpha - \phi) < \theta < \pi\phi.$$

This particular situation is valid if $\phi + \alpha < 1$.

To find the disturbance in the elliptic region, we use the conformal transformation $\zeta = \operatorname{sech}(\nu + i\theta)/\phi$, which maps the sector of the circle $\lambda < c$, $0 < \theta < \phi\pi$ into the upper half ζ plane. The conditions which must be satisfied by the function $\partial\omega/\partial\zeta$ are exactly those in the half-plane diffraction problem; and indeed, (10) gives the formula for $\partial\omega/\partial\zeta$ in the wedge problem if we take $\beta = 2(1 - \alpha - \phi)/\phi$ and $\gamma = 2(1 + \alpha - \phi)/\phi$. Integration of (10) in this case also leads us to (5.5.7) of Friedlander, [3].

It remains to be emphasized that this technique and the conical flow method of Buseman (see, e.g., [7]) are very closely related. Keller and Blank [6] described the results of this method very fully for all possible situations in the problem of pulse diffraction by a wedge; however their method cannot be extended to provide the solution of problems complicated by refraction effects.

¹ See Ch. 5 of [3].

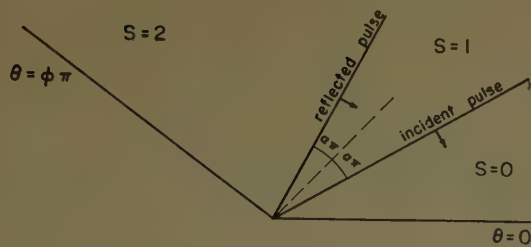
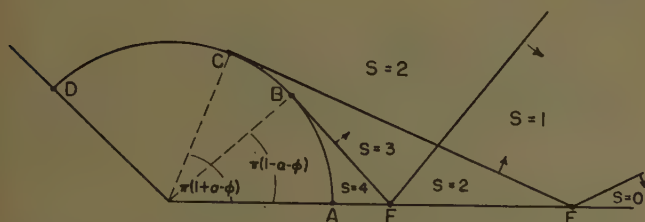


Fig. 3—The initial state in the wedge diffraction problem.

Fig. 4—The solution in the hyperbolic region of the (λ, θ) plane.

III. PULSE DIFFRACTION AND REFRACTION IN A COMPOSITE SYSTEM

A. Introduction

The value of the technique described in Section II does not lie in the fact that it provides yet another, slightly different means of solving the classical half-plane problem. Indeed, the only reason for the inclusion of the details in Section II is to confirm the value of the assumption of dynamic similarity in a diffraction problem.

We have already shown that in these two dimensional problems the problem of finding the vector solutions of Maxwell's equations reduces to that of finding solutions to the scalar wave equation with the boundary coordinates $s=0$ or $\partial s/\partial\theta=0$. This situation is exactly analogous to that in acoustics if we take s to represent the condensation, so that the fluid velocity \mathbf{q} is given by

$$\partial \mathbf{q} / \partial t = -c^2 \nabla s; \quad (12)$$

and hence the radial and transverse components of velocity (u, v) are given by

$$\lambda \frac{\partial u}{\partial \lambda} = c^2 \frac{\partial s}{\partial \lambda}, \quad \lambda^2 \frac{\partial v}{\partial \lambda} = c^2 \frac{\partial s}{\partial \theta}. \quad (13)$$

In this section the method described above is extended to give the solution in the case in which a plane pulse is diffracted by a half-plane which lies between two regions of different physical properties. The half-plane $\theta = \pm\pi$ is either perfectly reflecting or perfectly absorbent. The velocity of sound in the fluid of density ρ is c , and wherever qualification is necessary, we show which medium we are referring to by the suffixes 1 or 2.

There are several aspects of the physical problem which are of particular interest. The singularity at the apex of the half-plane is the only known result in the

steady-state problem. Meixner [5], who examined the power series expansion of the solution of the reduced-wave equation corresponding to a sinusoidal time dependence, found the edge singularity in this geometry to be unaffected by the discontinuity in physical properties.

The manner in which a disturbance passes from one medium into another has not been discussed before. Of course, the Fresnel coefficients for steady plane pulse reflexion and refraction at a plane surface are well-known, but these coefficients are without meaning when total reflexion occurs. Friedlander [4] found the form of the disturbance in this case, but since both these situations are essentially steady ones, they give no insight into the transient problem.

Craggs [1], [2], who has provided the greater part of the basis for the method of solution in this paper, examined a situation involving two right-angled wedges. No diffraction occurs in this case, and there is no singularity at the common apex; it appears that his method needs modifying when diffraction effects occur.

B. Half-Plane on the Surface Between Two Media

We shall first examine in detail the case in which the diffracting obstacle is a perfect reflecting half-plane ($r>0, \theta = \pm\pi$) lying on the plane of separation between two media. We consider the situation in which the velocity of sound c_1 , in medium 1 ($r>0, 0>\theta>-\pi$) is greater than that (c_2) in medium 2 ($r>0, 0<\theta<\pi$). We put $c_2 = mc_1$, and $\rho_2 = k\rho_1$.

The problem is to examine the propagation of a plane pulse in medium 2 which makes an angle ϕ with the normal to the half-plane. As in Section II, in order to make the assumption of dynamic similarity plausible, we must set up as an initial state one which satisfies all the appropriate boundary conditions. These conditions are that $\partial s_1/\partial\theta=0$ for $\theta=-\pi$, and that $\partial s_2/\partial\theta=0$ for $\theta=\pi$. The conditions at the common surface between the two media are found from the continuity of pressure and normal velocity. Thus for $r>0, \theta=0$, $s_1 = m^2 k s_2$ and $\partial s_1/\partial\theta = m^2 \partial s_2/\partial\theta$. In order to avoid having to set up a refracted disturbance before the pulse reaches the edge of the half-plane, we must restrict our consideration to the case in which the pulse arrives from the second quadrant of the (r, θ) plane; this pulse is accompanied by the reflected pulse. The initial state is depicted in Fig. 5. The assumption of dynamic similarity having been made, we must now find suitable solutions for (3) and (4).

C. The Solution in the Hyperbolic Region

The solution in the greater part of the hyperbolic region is determined by the values of s as $\lambda \rightarrow \infty$; that is, by the initial conditions of the problem. It is found by the method of characteristics. There are two distinct situations, according to whether $\phi <$ or $> \psi$, where $\psi = \sec^{-1}(1/m)$ is the critical angle; for $\phi < \psi$ total re-

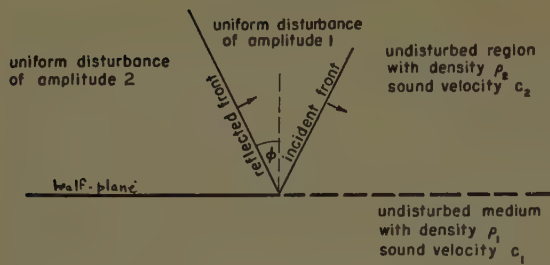


Fig. 5—The initial state in the diffraction and refraction of a pulse by a perfectly reflecting half-plane.

flexion occurs. In the (λ, θ) plane the hyperbolic region lies outside the sonic circle $\lambda = c_1$ in medium 1, and outside the sonic circle $\lambda = c_2$ in medium 2; the values of s in this region are shown in Figs. 6 and 7. The initial fronts are represented by the lines DI and ER . DE is the front of the pulse reflected at the refracting surface, and when the point D is outside the larger sonic circle, DT is the front of the refracted pulse. The line BC , which is a characteristic line, is the front of the disturbance which having entered medium 1 moves along the interface at grazing incidence and is refracted back into medium 2. The solution at this stage is fully determined in the whole region outside the sonic circles, except within the triangle ABC . Inside this region we introduce the unknown function $g(\mu_2 + \theta)$ to represent the part of s affected by the disturbance in the elliptic region of medium 1. [This is the function g of (3), with the function f a constant.] The constants R and T in the normally refracting case (Fig. 7) are the Fresnel coefficients given by

$$R = \frac{\tan \phi - k(m^2 \sec^2 \phi - 1)^{1/2}}{\tan \phi + k(m^2 \sec^2 \phi - 1)^{1/2}},$$

$$T = \frac{2m^2 k \tan \phi}{\tan \phi + k(m^2 \sec^2 \phi - 1)^{1/2}}. \quad (14)$$

From Fig. 6 we see that on $\lambda = c_2$ in medium 2

$$s_2 = 1 + g \quad \text{for } 0 < \theta < \phi,$$

$$s_2 = 2 + g \quad \text{for } \phi < \theta < \psi,$$

$$s_2 = 2 \quad \text{for } \psi < \theta < \pi,$$

while $s = 0$ on $\lambda = c_1$ in medium 1. From Fig. 7 we see that on $\lambda = c_2$ in medium 2

$$s_2 = R + g \quad \text{for } 0 < \theta < \psi,$$

$$s_2 = 1 + R \quad \text{for } \psi < \theta < \phi,$$

$$s_2 = 2 \quad \text{for } \phi < \theta < \pi;$$

while on $\lambda = c_1$ in medium 1

$$s_1 = T \quad \text{for } 0 > \theta > -\chi,$$

$$s_1 = 0 \quad \text{for } -\chi > \theta > -\pi,$$

where $\chi = \sec^{-1} m \sec \phi$.

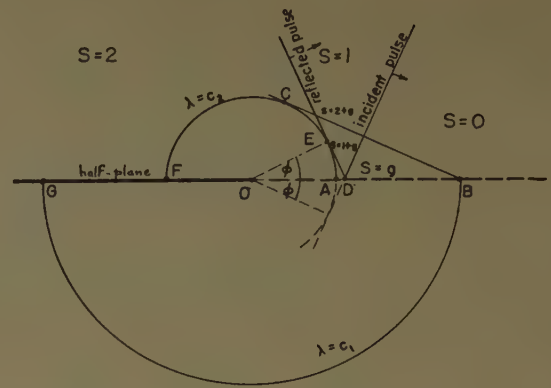


Fig. 6—The solution in the hyperbolic region of the (λ, θ) plane with $c_2 > c_1$ for sub-critical incidence $c_2 \sec \phi < c_1$.

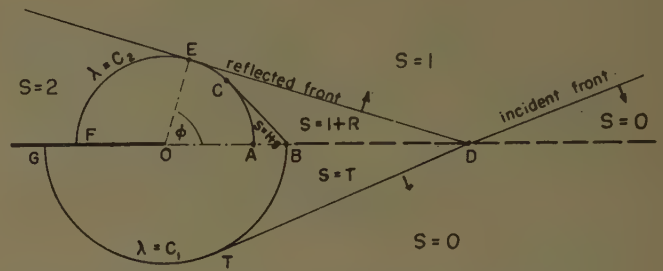


Fig. 7—The solution in the region of the (λ, θ) plane with $c > c$ for supercritical incidence $c_2 \cos \theta < c_1$.

D. The Elliptic Region of the (λ, θ) Plane

We use the conformal transformation

$$\zeta_1 = \xi_1 + i\eta_1 = \operatorname{sech}(\nu_1 + i\theta)$$

to map the inside of the semicircle $\lambda = c_1$, $-\pi < \theta < 0$ into the lower half of the complex ζ_1 plane, and the transformation

$$\zeta_2 = \xi_2 + i\eta_2 = \operatorname{sech}(\nu_2 + i\theta)$$

to map the inside of the semicircle $\lambda = c_2$, $0 < \theta < \pi$ into the upper half of the complex ζ_2 plane. The points F, O, A, E , and C in medium 2 are mapped into the points $\zeta_2 = -1, 0, 1, \sec \phi$, and $1/m$, respectively, while the points G, O, A, D , (or T), and B in medium 1 are mapped into the points $\zeta_1 = -1, 0, m, m \sec \phi$, and 1 .

Consider first the solution in the lower half of the ζ_1 plane. Since s is piecewise constant on the semicircle $\lambda = c_1$, it follows as in Section III-B that

- 1) $\partial \omega_1 / \partial \zeta_1$ is imaginary on the section of the real axis $|\xi_1| > 1$.

From the condition that $\partial s / \partial \theta = 0$ on OG , it follows that

- 2) $\partial \omega_1 / \partial \zeta_1$ is real on the segment of the real axis $0 < \xi_1 < 1$.

On the upper side of the line AB we know that the solution takes the form $s = g(\mu_2 + \theta) + \text{constant}$. It follows from this solution and from the condition of continuity of pressure and normal velocity that

$$\begin{aligned}\frac{\partial s_1}{\partial \lambda} &= m^2 k \frac{\partial s_2}{\partial \lambda} = m^2 k \frac{\partial \mu_2}{\partial \lambda} \frac{\partial s_2}{\partial \theta} = k \frac{\partial \mu_2}{\partial \lambda} \frac{\partial s_1}{\partial \theta} \\ &= -k \frac{\partial \mu_2}{\partial \lambda} \frac{\partial \tau_1}{\partial \nu_1} = -k \frac{\partial \mu_2}{\partial \nu_2} \frac{\partial \tau_1}{\partial \lambda}.\end{aligned}$$

Hence, we have the condition that

$$\frac{\partial s_1}{\partial \lambda} + k \frac{\partial \mu_2}{\partial \nu_1} \frac{\partial \tau_1}{\partial \lambda} = 0;$$

i.e., that

$$3) \quad Rl \left[\frac{\partial \omega_1}{\partial \lambda} \left(1 - ik \frac{\partial \mu_2}{\partial \nu_1} \right) \right] = 0.$$

Thus, if $F(\xi_1)$ is a real valued function on the segment $m < \xi_1 < 1$, then on this segment

$$\frac{\partial \omega_1}{\partial \xi_1} = F(\xi_1)/1 + \frac{i}{mk} \left(\frac{\xi_1^2 - m^2}{1 - \xi_1^2} \right)^{1/2}. \quad (15)$$

The other conditions to be imposed, as in Section III-B, are that

- 4) the only singularities of $\partial \omega_1 / \partial \xi_1$ are at points $\xi_1 = \pm 1, +m, 0$, as well as at the points $\xi_1 = m \sec \phi$,
- 5) no singularity of $\partial \omega_1 / \partial \xi_1$ may be of higher order than a simple pole,
- 6) there is an edge condition $\partial \omega / \partial \xi = 0(\xi^\delta)$ with $\delta > 1$ as $\xi \rightarrow 0$, and
- 7) the point at infinity is an ordinary point.

It is of interest to examine the function in the denominator in (15). This function is regular everywhere in the complex ξ_1 plane except on the real axis. It has branch points at $\xi_1 = \pm m, \pm 1$, and the branches of the radicals are chosen so that it has no zeros. The singularity at $\xi_1 = -m$ is not one of those allowed under Condition 4), so we must factorize this function in a manner which enables us to remove this singularity from the solution without changing its complex form on the real axis, for $m < \xi_1 < 1$.

E. The Factorization of the Complex Function

Consider the complex function

$$M(\xi) = mk + i \left(\frac{\xi^2 - m^2}{1 - \xi^2} \right)^{1/2} / 1 + mk$$

which is the analytic continuation of the function defined on the segment of the real axis $m < \xi < 1$. The properties of $L(\xi) = \ln M(\xi)$ are that

- 1) as $|\xi| \rightarrow \infty$, $L(\xi) = 0(\xi^{-2})$;
- 2) $L(\xi)$ is regular in the infinite strip $m < Rl(\xi) < m$;
- 3) $L(\xi)$ is a real function of ξ within the same strip, since the continuation of $L(\xi)$ is

$$\ln \left\{ mk + \left(\frac{m^2 - \xi^2}{1 - \xi^2} \right)^{1/2} / 1 + mk \right\};$$

and

- 4) $L(\xi)$ is single valued in the cut plane with one cut joining the points $-m, -1$ and another the points $+m, +1$.

If we take a rectangular contour C as shown in Fig. 8,

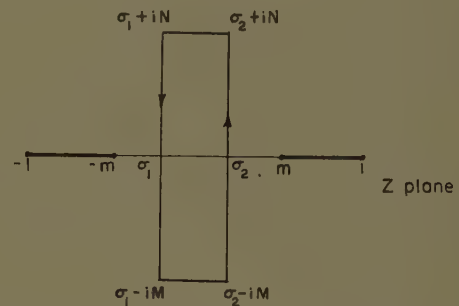


Fig. 8—The contour C of integration in the complex Z plane.

with sides joining the points $\sigma_1 + iN, \sigma_2 + iN, \sigma_1 - iM$, and $\sigma_2 - iM$, then for $-m < \sigma_1, \leq Rl(z) \leq \sigma_2 < m$, we are able to use Cauchy's integral to state that

$$L(\xi) = \frac{1}{2\pi i} \int_C \frac{L(z)}{z - \xi} dz.$$

Here the integrand is sufficiently small as $|z| \rightarrow \infty$ to allow the limiting process, $N \rightarrow \infty, M \rightarrow \infty$, to be carried out separately at the upper and the lower end of the rectangle. The contribution from the ends of the rectangle is vanishingly small, so we may replace the contour integral by the sum of two line integrals

$$\begin{aligned}L(\xi) &= \frac{1}{2\pi i} \int_{\sigma_2 - i\infty}^{\sigma_2 + i\infty} \frac{L(z)}{z - \xi} dz - \frac{1}{2\pi i} \int_{\sigma_1 - i\infty}^{\sigma_1 + i\infty} \frac{L(z)}{z - \xi} dz \\ &= L_-(\xi) - L_+(\xi).\end{aligned}$$

Each line integral $L_-(\xi)$ and $L_+(\xi)$ is uniformly convergent for values of ξ taken within the infinite rectangle $|Rl(\xi)| < m$. Taken separately each integral may be continued analytically into a complete half-plane, since $L_-(\xi)$ is regular for $Rl(\xi) < m$, and $L_+(\xi)$ is regular for $Rl(\xi) > -m$.

Now since

$$\begin{aligned}L_+(\xi) &= \frac{1}{2\pi i} \int_{\sigma_1 - i\infty}^{\sigma_1 + i\infty} \frac{L(z)}{z - \xi} dz \\ &= \frac{1}{2\pi} \int_0^\infty \left[\frac{L(\sigma_1 + i\eta)}{\sigma_1 + i\eta - \xi} + \frac{L(\sigma_1 - i\eta)}{\sigma_1 - i\eta - \xi} \right] d\eta,\end{aligned}$$

and since in $|z| < m$, $L(z)$ is a real function of z , for real values of ξ the integrand is the sum of complex conjugates, and the integral is therefore real for any real value of $\xi > -m$. Likewise $L(\xi)$ is real on the real axis for $\xi < m$.

We may therefore write

$$\exp - L_-(\zeta) = \exp - L_+(\zeta)/M(\zeta), \quad (16)$$

and we see that the function on either side of this equation is

- 1) regular in the half-plane $Re(\zeta) < m$,
- 2) real on the real axis in $Re(\zeta) < m$,
- 3) has the branch points and the complex behavior of $[M(\zeta)]^{-1}$ in the half-plane $Re(\zeta) \geq m$.

It is also easy to show that $L_+(-\zeta) = -L_-(\zeta)$; this is a consequence of the even nature of the function $L(z)$.

In its present form, the integral $L_+(\zeta)$ is not very suitable for computation. However, we may add to this line integral one for which the integral is taken round a semicircle of infinite radius. The path of integration is thus closed without changing the value of the integral. The integrand has branch points at $z = -m$ and $z = -1$, within the contour. The path of integration may now be deformed into a loop encircling the cut, so we find that

$$\pi L_+(\zeta) = \int_m^1 \tan^{-1} [(z^2 - m^2)^{1/2}/mk(1 - z^2)^{1/2}] dz/z + \zeta. \quad (17)$$

This integrand may be expanded in an ascending series of powers of $(m + \zeta)^{-1}$; the integration of the coefficients of this expansion in ascending powers of $(1 - m)$ is then easily carried out.

F. The Solution in the Elliptic Region

The expression which represents the solution to the problem and which satisfies the conditions in Section III-D may now be written down. The most general is that

$$\pi \frac{\partial \omega_1}{\partial \zeta_1} = \frac{\exp - L_+(\zeta_1)}{mk(1 - \zeta_1^2)^{1/2} + i(\zeta_1^2 - m^2)^{1/2}} \cdot \left\{ \frac{A + B \left(\frac{\zeta_1 - m}{\zeta_1} \right)^{1/2}}{(\zeta_1 - m \sec \phi)} \right\}, \quad (18)$$

where the function $L_+(\zeta_1)$ has been defined in Section III-E. The problem which remains is to find the disturbance in the elliptic region $\lambda < c_2$, $0 < \theta < \pi$. Now (18) has been derived for general values of ζ_1 in the lower half-plane by means of the principle of analytic continuation, and we may in turn continue this solution into the upper half ζ_2 plane, because the solutions in these two regions are linked across the line OA , on which $\zeta_1 = m\zeta_2$.

Across this line OA the continuity conditions may be written in the form

$$m^2 k \frac{\partial s_2}{\partial \zeta_2} = \frac{\partial s_1}{\partial \zeta_2}; \quad \frac{\partial \tau_1}{\partial \zeta_2} = m^2 \left(\frac{1 - \zeta_2^2}{1 - m^2 \zeta_2^2} \right)^{1/2} \frac{\partial \tau_2}{\partial \zeta_2}. \quad (19)$$

Since on the corresponding section of the real axis $0 < \zeta_2 < m$,

$$\pi \frac{\partial \omega_1}{\partial \zeta_1} = \frac{\exp - L_+(\zeta_1)}{mk(1 - \zeta_1^2)^{1/2} + (m^2 - \zeta_1^2)^{1/2}} \cdot \left\{ \frac{A - iB(m - \zeta_1/\zeta_1)^{1/2}}{(\zeta_1 - m \sec \phi)} \right\}, \quad (20)$$

it follows from (19) and (20) that

$$m^2 k \pi \frac{\partial \omega_2}{\partial \zeta_2} = \frac{\exp - L_+(m\zeta_2)}{mk(1 - m^2 \zeta_2^2)^{1/2} + m(1 - \zeta_2^2)^{1/2}} \cdot \left\{ \frac{A - ikB \left(\frac{1 - m^2 \zeta_2^2}{\zeta_2(1 + \zeta_2)} \right)^{1/2}}{(\zeta_2 - \sec \phi)} \right\} \quad (21)$$

on the section of the real axis $0 < \zeta_2 < 1$.

The function

$$\left\{ 1 + \frac{1}{k} \left(\frac{1 - \zeta_2^2}{1 - m^2 \zeta_2^2} \right)^{1/2} \right\} \exp + L_+(m\zeta_2)$$

is regular in the half-plane $Re(\zeta_2) < 1$; it is real on the real axis in this domain of regularity; and it has the complex behavior and the branch points of the function

$$1 + \frac{i}{k} \left(\frac{\zeta_2^2 - 1}{1 - m^2 \zeta_2^2} \right)^{1/2}$$

in the half-plane $Re(\zeta_2) \geq -1$.

As in the case of the other elliptic region we expect the function $\partial \omega_2 / \partial \zeta_2$ to have the properties that

- 1) it has singularities at $\zeta_2 = 0, \pm 1, +1/m$, and $+ \sec \phi$;
- 2) the point at infinity is an ordinary point;
- 3) it is imaginary on the real ζ_2 axis when $\zeta_2 > 1/m$, $\zeta_2 < -1$, and it is real on the real ζ_2 axis when $-1 < \zeta_2 < 0$.

There is, however, a branch point at $\zeta_2 = -1/m$ in the expression on the right-hand side of (21). The only possible way in which (18), which has the correct behavior in the lower half-plane, may be continued into an expression correct in the upper region is to take $A = 0$; there is, therefore, only the constant B which remains to be found.

G. The Determination of the Constant B

It will be remembered that in Figs. 4 and 5, we have distinct representations of the problem for the cases in which $m \sec \phi < 1$ and in which $m \sec \phi > 1$. In the case of supercritical incidence shown in Fig. 5, we may use the method of residues to relate the discontinuity on the sonic circle either at the point T or at the point E ; we find the same result in each case. It follows that

The quantities T and R are the Fresnel coefficients defined previously in (14).

The triangle ABC , which results from the refraction of the disturbance in the elliptic region of medium 2 which travels at grazing incidence, is now in medium 1. Now the conditions to be imposed on $\partial\omega_2/\partial\zeta_2$ across the line AB is that

$$\partial\omega_2/\partial\zeta_2 = P(\zeta_2)/1 - \frac{i}{k} \left(\frac{1 - \zeta_2^2}{m^2\zeta_2^2 - 1} \right)^{1/2},$$

where $P(\zeta_2)$ is a function which takes real values on the real ζ_2 axis when $1/m < Rl(\zeta_2) < 1$. To remove the unwanted branch points at the points $\zeta_2 = -1/m, -1$, we must introduce the factor $\exp +M(m\zeta_2)$, where

$$\pi M_+(\zeta) = \int_1^m \tan^{-1} \left\{ \frac{(m^2 - z^2)}{mk(z^2 - 1)} \right\}^{1/2} \frac{dz}{z + \zeta}. \quad (31)$$

This function is derived exactly as in Section III-E and its properties are very similar. We may now write down an expression for $\partial\omega_2/\partial\zeta_2$ which satisfies all the necessary conditions, and as before, we are only left with one constant to be determined. Thus

$$m^2\pi \frac{\partial\omega_2}{\partial\zeta_2} = \frac{B \exp M_+(m\zeta_2)}{m \left[1 - \frac{i}{k} \left(\frac{1 - \zeta_2^2}{m^2\zeta_2^2 - 1} \right)^{1/2} \right]} \cdot \frac{\left(\frac{m\zeta_2 - 1}{m\zeta_2} \right)^{1/2}}{(1 - \zeta_2^2)^{1/2}(\zeta_2 - \sec \phi)}. \quad (32)$$

It is necessary to make $\partial\omega_2/\partial\zeta_2$ imaginary on the real ζ_2 axis when $0 < \zeta_2 < 1/m$ to avoid introducing branch points which give an incorrect solution. On applying the continuity conditions across OA we find that

$$m\pi \frac{\partial\omega_1}{\partial\zeta_1} = - \frac{iB \exp M_+(\zeta_1)(1 - \zeta_1/\zeta_1)^{1/2}}{\left[1 + \frac{1}{mk} \left(\frac{m^2 - \zeta_1^2}{1 - \zeta_1^2} \right)^{1/2} \right] (1 - \zeta_1^2)^{1/2}(\zeta_1 - m \sec \phi)}, \quad (33)$$

and this expression satisfies all the required conditions.

The constant B , which is found by the method of residues, is given by

$$B = 2m^3 \exp$$

$$- M_+(m \sec \phi)(\sec \phi - 1) \tan \phi \left[\frac{m \sec \phi + 1}{m \sec \phi - 1} \right]^{1/2}. \quad (34)$$

This completes the description of the calculation of the velocity and condensation (or pressure) derivatives in the acoustic problem, when the angle of incidence, as shown in Fig. 5, lies in the range $0 < \phi < \pi/2$. The results are directly applicable to the electromagnetic problem in the magnetic mode, with a perfectly conducting half-plane, provided that we take the constant k to define the ratio of the dielectric constants of the two media instead of the density ratio.

I. The Perfectly Soft Half-Plane

When we consider the problem in which s vanishes at the surface of the half-plane, there is only a slight change in the method of solution. To satisfy this boundary condition the reflected pulse must annul instead of reinforce the incident pulse. The initial situation must be changed accordingly, as in Fig. 10. For the solution in the hyperbolic region in the various cases we may refer to Figs. 6, 7, and 9, if we reduce the amplitude of the disturbance behind the front ER by 2.

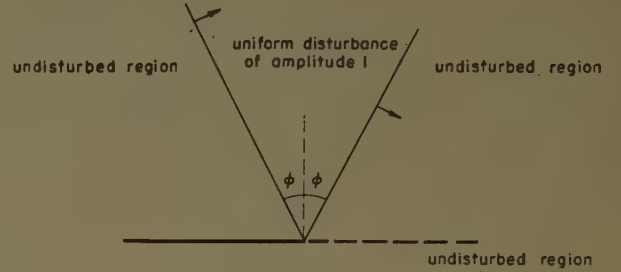


Fig. 10—The initial state for a perfectly absorbing half-plane.

Within the elliptic regions the vanishing of s on $\theta = \pm\pi$ means that $\partial s/\partial\lambda = 0$ on the line $\lambda < c$, $\theta = \pm\pi$. Hence $\partial\omega/\partial\zeta$ must be imaginary on the real axis for $-1 < \zeta_1 < 0$. All the conditions of the problem may be satisfied by

$$\pi \frac{\partial\omega_1}{\partial\zeta_1} = \frac{A \exp - L_+(\zeta_1) \left(\frac{1 - \zeta_1}{\zeta_1} \right)^{1/2}}{\left[1 + \frac{i}{mk} \left(\frac{\zeta_1^2 - m^2}{1 - \zeta_1^2} \right)^{1/2} \right] (\zeta_1 - m \sec \phi)}, \quad (35)$$

and

$$m^2k\pi \frac{\partial\omega_2}{\partial\zeta_2} = \frac{A \exp - L_+(m\zeta_2) \left(\frac{1 - m\zeta_2}{m\zeta_2} \right)^{1/2}}{1 + \frac{1}{k} \left(\frac{1 - \zeta_2^2}{1 - m^2\zeta_2^2} \right)^{1/2} (\zeta_2 - \sec \phi)}. \quad (36)$$

No solution which has an imaginary part on the interface between the two elliptic regions is correct, since it must then, according to the continuity conditions (19), contain unwanted branch points in one medium or the other.

When $m \sec \phi > 1$, the constant A must take the value

$$2m^2 \exp L_+(m \sec \phi) \tan \phi \left(\frac{m \sec \phi}{m \sec \phi + 1} \right)^{1/2} (m \sec \phi - 1).$$

When $m \sec \phi > 1$, total reflexion is expected, and when we use the method of residues to find the constant A and its relation to the jump across the front reflected on the interface, we find that this value for A is unchanged.

When $m > 1$, as in Fig. 9, the results in the elliptic regions are that

$$m^3 \pi \partial \omega_2 / \partial \xi_2$$

$$= \frac{A \exp M_+(m \xi_2) \left(\frac{1 - \xi_2^2}{\xi_2} \right)^{1/2}}{\left[1 - \frac{i}{k} \left(\frac{1 - \xi_2^2}{m^2 \xi_2^2 - 1} \right)^{1/2} \right] (\xi_2 \sec \phi)} \quad (37)$$

and

$$\pi \partial \omega_1 / \partial \xi_1$$

$$= \frac{A \exp M_+(\xi_1) \left(\frac{m - \xi_1}{\xi_1} \right)^{1/2}}{\left[1 + \frac{1}{mk} \left(\frac{m^2 - \xi_1^2}{1 - \xi_1^2} \right)^{1/2} \right] (\xi_1 - m \sec \phi)} \quad (38)$$

Here A must have the value

$$2m^3 \tan \phi \left[\frac{\sec \phi (\sec \phi - 1)}{m^2 \sec^2 \phi - 1} \right]^{1/2} \frac{\exp - M_+(m \sec \phi)}{\sec \phi - 1} \quad (39)$$

These results may also be used in the electromagnetic problem in the electric mode, with a perfectly conducting half-plane, if we take s to be the quantity $E_z/\mu c^2$, and if we replace the density ratio k by the ratio of the permeabilities μ of the two media.

IV. CONCLUSION

By assuming dynamic similarity in the solution, we have found explicit formulas for the derivatives of the field components in a series of pulse diffraction problems. These results show clearly how the pulses are propagated and also how a refracted disturbance is set up even in the case of total reflexion. A most interesting result is that the dynamically similar solution for the problem of diffraction by a half-plane $r > 0$, $\theta = \pm \pi$ has the same form of solution on the half-plane $\theta = 0$ within the smaller sonic circle, whether the media in the two regions $0 > \theta > -\pi$ and $0 < \theta < \pi$ have distinct properties or not. This statement is the immediate consequence of the fact that $\partial \omega / \partial \xi$ is either real in the absorbent half-plane problem or purely imaginary in the other case. The results in the homogeneous medium follow directly from (11) and (13). Thus, for the reflecting half-plane the pressure is constant, and for the absorbing half-plane the normal velocity vanishes on a steadily expanding section of the half-plane $\theta = 0$.

The method described may also be used to analyze the case in which the incident pulse passes into a medium in which the velocity of propagation depends on the direction of motion of the disturbance.

BIBLIOGRAPHY

- [1] J. W. Craggs, *Proc. Roy. Soc. (London) A.*, vol. 237, p. 372; 1956.
- [2] J. W. Craggs, *J. Fluid Mech.*, vol. 3, p. 176; 1957.
- [3] F. G. Friedlander, "Sound Impulses," Cambridge University Press, Cambridge, Eng.; 1958.
- [4] F. G. Friedlander, *Quart. J. Mech.*, vol. 4, p. 344; 1948.
- [5] J. Meixner, *Inst. Math. Sciences, New York Univ., New York, N. Y.*, Res. Rept. E M 72; 1954.
- [6] J. B. Keller and A. Blank, *Commun. Pure Applied Math.*, vol. 4, p. 75; 1951.
- [7] G. N. Ward, "Linearized Theory of Steady High Speed Flow," Cambridge University Press, Cambridge, Eng.; 1955.

The Field of a Pulsed Dipole in an Interface

C. S. GARDNER† AND J. B. KELLER†

ABSTRACT

A VERTICAL electric dipole in the plane interface between two semi-infinite homogeneous dielectric media is switched on at $t=0$. The resulting electromagnetic field is determined by a new

method based on a new representation of the field of a point source as the superposition of the fields of line sources. The field of a line dipole is found by using, first, the conical flow transformation, and then a Jonkowski transformation which leads to a function-theoretic problem of the Riemann-Hilbert type. This problem is solved explicitly and an explicit expression is obtained for the point dipole field at points of the interface. The results are finally compared with those of others.

† Institute of Mathematical Sciences, New York University, New York, N. Y.

Diffraction by a Half-Plane with a Special Impedance Variation*

J. SHMOYS†

Summary—This paper deals with the diffraction of a plane electromagnetic wave incident on a variable impedance half-plane, at right angles to the edge. The problem is dealt with by a method closely related to separation of variables in parabolic coordinates. A class of solutions of the wave equation is obtained in this manner; the boundary conditions which may then be imposed on the half-plane are investigated. These boundary conditions can be chosen in such a way as to permit arbitrary values of reflection and transmission coefficients for the half-plane far away from the edge. The dependence of the diffraction pattern on these coefficients will be discussed.

THIS investigation deals with the diffraction of a plane wave by a variable impedance half-plane. The purpose of the study is to find some impedance variation for which an exact solution of the diffraction problem exists, and then ascertain whether this solution can be interpreted in terms of geometrical optics. The solution found is a slight generalization of that used by Raman and Krishnan.¹

The method used is based on the fact, discovered by Lamb² and exploited more fully by Hanson,³ Brillouin,⁴ Bouwkamp⁵ and Karp,⁶ that the wave equation can be solved by separation of variables in a mixed coordinate system consisting of parabolic cylinder and cartesian coordinates. If we look for solutions of the wave equation in two dimensions:

$$\nabla^2 u + k^2 u = 0 \quad (1)$$

of the form

$$u(x) = e^{ik \cdot r} f(a\xi + \beta\eta), \quad (2)$$

where ξ and η are parabolic cylinder coordinates (Fig. 1),

$$\xi = \sqrt{2r} \sin(\theta/2) \quad (3)$$

$$\eta = \sqrt{2r} \cos(\theta/2), \quad (4)$$

then we find that there are two solutions for every direction of the vector k . One of these, however, contains an

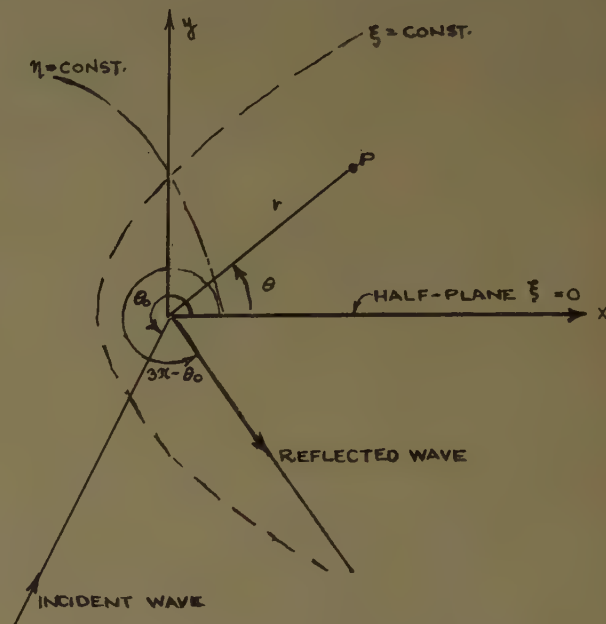


Fig. 1—Physical configuration and choice of coordinates.

incoming cylindrical wave and must therefore be discarded. For the remaining solution

$$\alpha/\beta = \tan(\theta_0/2), \quad (5)$$

where θ_0 is the direction of k

$$k_y/k_x = \tan \theta_0.$$

The function f in (2) is a Fresnel integral. The solution of a diffraction problem must consist of two terms of form (2), one in which $k = k_{\text{inc}}$, and one in which $k = k_{\text{refl}}$. The most general solution of the wave equation containing the appropriate plane waves, and only an outgoing cylindrical wave, is then

$$u(r) = \exp ik(-x \cos \theta_0 - y \sin \theta_0)$$

$$\begin{aligned} & \cdot \left[f_0 + f_1 \int_{\sqrt{2r} \cos(\theta-\theta_0)/2}^{\infty} e^{ikt^2} dt \right] \\ & + \exp ik(-x \cos \theta_0 + y \sin \theta_0) \\ & \cdot \left[g_0 + g_1 \int_{\sqrt{2r} \cos(\theta+\theta_0)/2}^{\infty} e^{ikt^2} dt \right]. \quad (6) \end{aligned}$$

The four arbitrary constants in this solution are simply related to the amplitudes of incoming and outgoing plane waves contained in the solution. Thus, if the incident wave amplitude is 1, the reflected wave amplitude is R (see Fig. 2), the transmitted wave amplitude is T

* The work done in this paper was sponsored by the AF Cambridge Research Center, under Contract No. AF-19(604)-4143, and M.I.T. Lincoln Lab., Cambridge, Mass., Subcontract No. 96 under Contract AF-19(122)-458, PO. No. 1096.

† Microwave Research Inst., Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

¹ C. V. Raman and K. S. Krishnan, "The diffraction of light by metallic screen," *Proc. Roy. Soc. (London)*, vol. 116A, pp. 254-267; 1927.

² H. Lamb, "Hydrodynamics," Dover Publications, Inc., New York, N. Y., pp. 538-541; 1945.

³ E. T. Hanson, "Shadow of a straight edge," *Phil. Trans. Roy. Soc. (London)*, vol. A237, p. 35; 1938.

⁴ J. Brillouin, "Demonstration directe des formules de Fresnel pour la diffraction d'une onde plane par un demi-plan réfléchissant," *C. R. Acad. Sci. (Paris)*, vol. 229, p. 513; 1949.

⁵ C. J. Bouwkamp, "Diffraction Theory," Math. Res. Group, New York University, N. Y., Res. Rept. No. EM-50; April, 1953.

⁶ S. N. Karp, "Reflection and transmission by a class of curved dielectric layers," to be published.

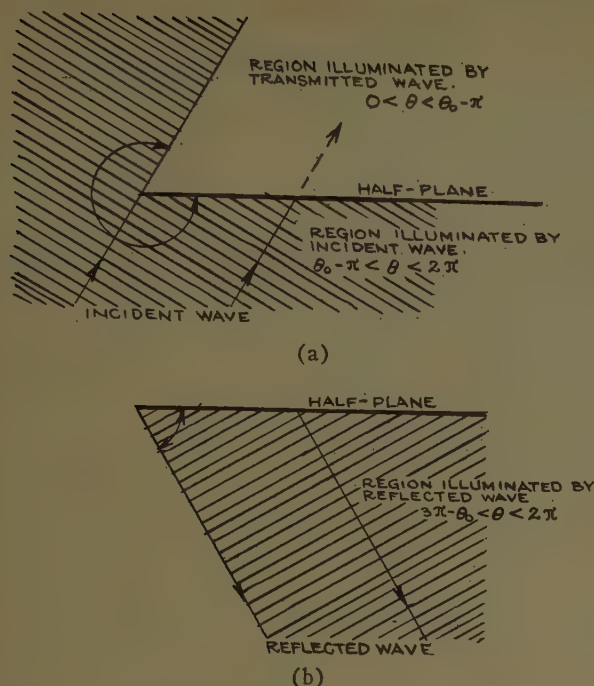


Fig. 2—Geometric-optical regions. (a) Incident wave domain. (b) Reflected wave domain.

and there is no incoming plane wave in the direction k_{refl} , then

$$\begin{aligned} f_0 &= 1 & f_1 &= (T - 1)/\sqrt{k/\pi i} \\ g_0 &= R & g_1 &= -R/\sqrt{k/\pi i}. \end{aligned} \quad (7)$$

This differs from Raman and Krishnan only in the added term due to transparency of the screen.

The diffracted field, or the cylindrical wave contained in (6) is given by

$$u_{\text{diff}} = \frac{e^{i(kr - 3\pi/4)}}{2\sqrt{2kr}} \cdot \left[(1 - T) \sec \frac{\theta - \theta_0}{2} + R \sec \frac{\theta + \theta_0}{2} \right]. \quad (8)$$

We see that slight variations in the reflection and transmission coefficients do not have a pronounced effect on the diffraction pattern.

A plot of the diffraction pattern for a wave incident normally on a perfectly absorbing half-plane with $R=0$, $T=0$ is shown in Fig. 3, together with the patterns for the perfectly reflecting half-planes, $R=\pm 1$, $T=0$, and the half-plane which absorbs perfectly all "angularly propagating" waves (Sommerfeld's "black screen"). For the latter, the expression in the square brackets in (8) is replaced by $4\pi/[\pi^2 - (\theta - \theta_0)^2]$. One notes the close agreement between the patterns for the two types of absorbing surfaces.

So long as the half-plane is not opaque, i.e., $T \neq 0$, the problem of diffraction by such a half-plane cannot be stated as a boundary value problem. Since we have been dealing with the scalar problem, it is necessary to identify the scalar u with a field component. In the trans-

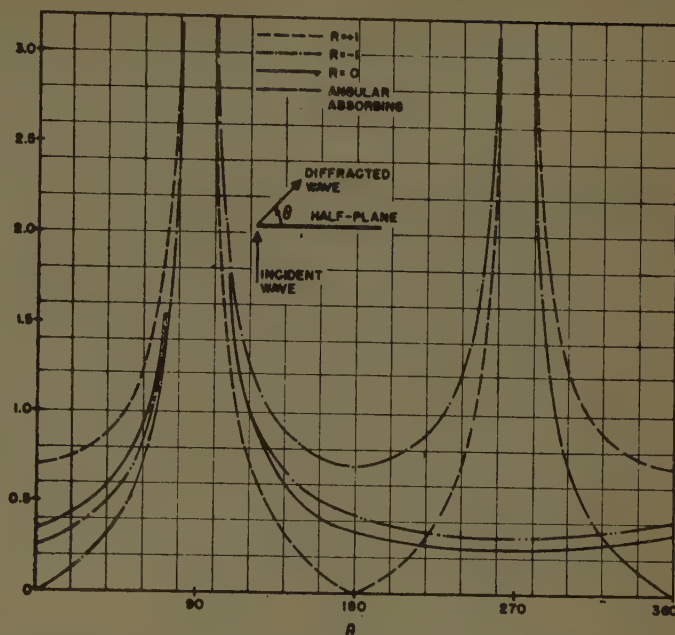


Fig. 3—Diffraction patterns of conducting and absorbing half-planes.

verse magnetic case (H parallel to the edge of the half-plane) $u = H_z$, and the normalized surface impedance is

$$Z_s^{\text{TM}} = \mp \frac{\partial u / \partial y}{iku} \Big|_{y=0} \quad (9)$$

where the upper or lower sign is to be used on the upper ($y=0+$) or lower ($y=0-$) side of the half-plane. In the transverse electric case (E parallel to the edge), $u = E_z$, and the normalized surface impedance is

$$Z_s^{\text{TE}} = \mp \frac{iku}{\partial u / \partial y} \Big|_{y=0} = (Z_s^{\text{TM}})^{-1}. \quad (10)$$

We can now interpret the solution given by (6) and (7) as a solution of a boundary value problem involving a half-plane with a certain variable surface impedance. The surface impedance is the following,

$$Z_s^{\text{TM}} = Y_s^{\text{TE}} = \left(\frac{1 - R}{1 + R} \right) \left[\pm \sin \theta_0 - \frac{\frac{1}{2} \sqrt{2kr} \sin \frac{\theta_0}{2} \exp \left(2ikr \cos^2 \frac{\theta_0}{2} \right)}{ikr \Phi \left(\pm \sqrt{2kr} \cos \frac{\theta_0}{2} \right)} \right]. \quad (11)$$

$$\Phi(z) = \int_{-\infty}^{\infty} \exp(i\xi^2 z) d\xi.$$

While this is the general result, valid for all angles of incidence, reflection coefficients, and distances from the edge, we wish to point out several properties of this impedance function. First, the distance dependence along the half-plane enters through the factor kr , so that the electrical radian is a natural unit of length. Second, the surface impedance tends to a constant very

rapidly as we move away from the edge. On the illuminated side, the surface impedance tends to

$$\frac{1-R}{1+R} |\sin \theta_0|, \quad (12)$$

the impedance of an infinite plane which would produce a reflection coefficient R . On the shadow side

$$Z_s^{\text{TM}} = Y_s^{\text{TM}} \rightarrow 0. \quad (13)$$

The variation of the impedance along the half-plane both on the illuminated and shadow sides is shown in Fig. 4 for the special $R=0$, *i.e.*, perfect absorption on the lit side. At the edge $Z_s^{\text{TM}} \rightarrow \infty$.

Finally, let us consider the special case $\theta_0 = \pi$. In this case the wave is incident along the half-plane, and the expression for the impedance reduces to

$$\begin{aligned} Z_s^{\text{TM}} = Y_s^{\text{TE}} &= \left(\frac{1-R}{1+R} \right) \frac{\frac{1}{2}\sqrt{2kr}}{ikr\Phi(0)} \\ &= \left(\frac{1-R}{1+R} \right) \frac{\sqrt{2}e^{-i3\pi/4}}{\sqrt{\pi kr}}. \end{aligned} \quad (14)$$

It is noted that the impedance varies in this case like $1/\sqrt{kr}$. The diffracted field is

$$u_{\text{diff}} = \frac{e^{i(kr+\pi/4)}}{2\sqrt{2kr}} (1+R) \sec \frac{\theta}{2}. \quad (15)$$

When we examine (14) and (15), we can see that the diffraction pattern is independent of the character of the surface, *i.e.*, the value of R . The impedance at the edge is infinite, far away from the edge it tends to zero. If the normalized impedance Z^{TM} , tends to zero so rapidly that one radian away from the edge it is already much smaller than 1, 0.1 for example (*i.e.*, $R \sim 1$), then the dif-

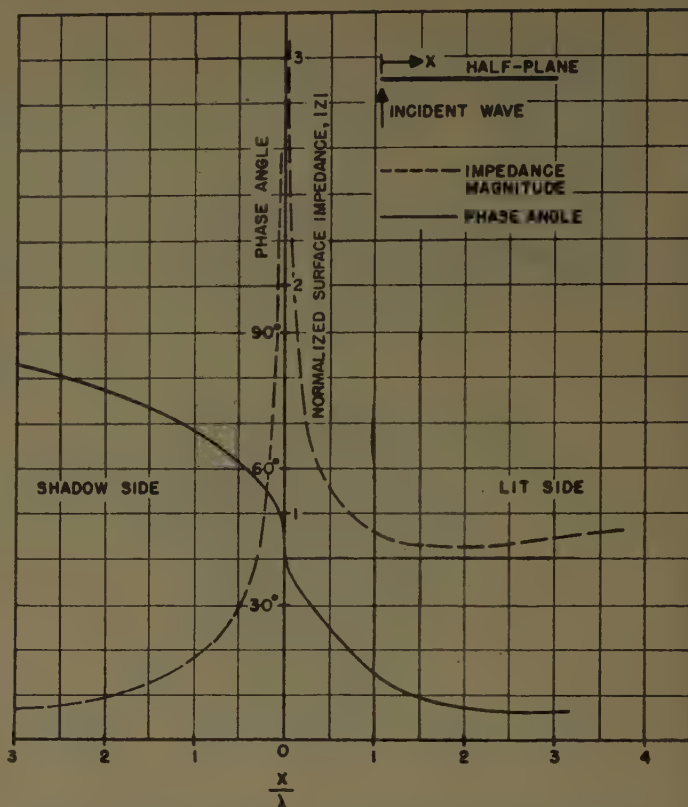


Fig. 4—Surface impedance variation on an absorbing half-plane.

fracted field is practically that (within 12 per cent) which would exist if the impedance were zero all the way to the edge. If, on the other hand the impedance drops off relatively slowly, *i.e.*, $R \approx -1$, so that at one radian it is still much greater than 1, then the diffracted field is almost the same as if the impedance of the half-plane were infinite.

A New Method for the Determination of Far Fields with Applications to the Problem of Radiation of a Line Source at the Tip of an Absorbing Wedge*

S. N. KARP† AND F. C. KARAL, JR.†

Summary—A new simple operational method for the determination of the diffracted far field for wedges with impedance boundary conditions is presented. The exterior angle of the wedge is $2\pi - \pi/2n$, where n is an integer. The excitation may be either an incident plane wave or a line source. The method is described in detail for the case of an absorbing wedge with a line source located at the tip. Several examples are given to illustrate the method. It is also shown that the method can be applied directly to the classical problem of perfectly conducting wedges.

I. INTRODUCTION

IN this paper we present a simple method for the determination of the diffracted far field for wedges with impedance boundary conditions. The excitation may be either an incident plane wave or a line source located at the tip of the wedge. The latter case is emphasized. The impedance boundary condition considered here is valid for the case of an absorbing surface. By an absorbing surface we mean a plane surface or screen that absorbs a maximum amount of radiation for certain angles of incidence of an incoming plane wave. The method to be described, however, is not restricted to absorbing surfaces but also applies to surfaces that support surface waves. Several examples are given to illustrate the method. It is also shown that the simple method developed here can be applied directly to the classical problem of diffraction by a perfectly conducting wedge or half plane.

The impedance boundary conditions prescribed on the wedge surfaces are given by

$$\frac{1}{r} \frac{\partial u}{\partial \theta} + \lambda u = 0, \quad \theta = 0, \quad (1)$$

$$\frac{\partial u}{\partial \theta} = 0, \quad \theta = \alpha, \quad (2)$$

where u is the z -component of the magnetic vector, r and θ are the usual polar coordinates, α is the angular space for which the solution is desired, and λ is a constant characteristic of the wedge surface.

When the surface is absorbing, the value of λ is given by

$$\lambda = iks, \quad (3)$$

where k is the free space propagation constant and s is a real positive parameter characteristic of the absorbing material. If the parameter s is restricted to lie between

0 and 1, the screen absorbs all radiation for some particular direction of incidence and partially absorbs radiation for other angles of incidence. Hence the absorbing surface under discussion here can act like a black screen for certain directions of incidence. For additional information on absorbing surfaces see Karal and Karp [2], and for information on diffraction by black screens see Baker and Copson [1].

The problems treated here are not separable because of the mixed boundary conditions. It is possible, however, to introduce an operator L which defines an auxiliary function $Lu \equiv v$, such that the auxiliary function satisfies the wave equation and simple homogeneous boundary conditions on both wedge surfaces. Once the auxiliary function is found, and this is not difficult, the original field could be determined by solving a partial differential equation. The boundary conditions in the original problem and certain necessary continuity conditions are then imposed, allowing one to obtain an exact formal solution valid for the entire region of physical space. This basic idea is due to Stoker [9] and Lewy [8] and has been adapted by the authors to the solution of problems in diffraction theory. (See [2]–[6].) However, if only the radiated far field is desired, it is possible to proceed in a simpler and more direct manner which completely avoids the detailed formal analysis which is quite lengthy and complicated. It is this simple and powerful method that is one of the main results of this paper. The other new results relate to the excitation of an absorbing wedge by a line source at the tip.

In Section II we determine the operator L that transforms the original impedance boundary conditions into simpler boundary conditions involving an auxiliary function whose solution is known. In Section III we make use of this operator and examine the problem of an incident plane wave striking a wedge of angle $\pi/2n$ (exterior angle $2\pi - \pi/2n$), where n is an integer, when an impedance boundary condition is prescribed on the upper wedge surface and a Neumann condition is prescribed on the lower wedge surface. The radiated far field is determined as a quotient of two trigonometric expressions. The numerator contains a set of unknown constants. The denominator has a corresponding number of zeros which would lead to undesirable shadow lines in the pattern. The constants in the numerator are then so determined as to eliminate these inadmissible poles or shadow lines.

The principal section is Section IV. In that section we determine the radiated far field of a line source located at the tip of a wedge of exterior angle $2\pi - \pi/2n$

* The research reported in this document has been sponsored by the AF Cambridge Research Center, Air Res. and Dev. Command, under Contract No. AF 19(604)5238.

† Institute of Mathematical Sciences, New York University, New York, N. Y.

when an impedance condition is prescribed on the upper face, and a Neumann condition is prescribed on the lower face.

Section V is devoted to illustrative examples of the general result of Section IV.

Section VI demonstrates how the method we have described can be employed in the direct derivation of the diffraction pattern for *perfectly* conducting wedges. In order to illustrate a certain alternative which is available in the application of the preceding method, we determine various constants in the far-field amplitude for diffraction by a perfectly conducting half plane and for a right-angled wedge by the use of the reciprocity theorem rather than by the method of regularity (*i.e.*, avoidance of undesirable poles or shadow lines) which we exploited in the preceding sections.

II. DETERMINATION OF OPERATOR

In the problems that follow we shall need an operator L that transforms the impedance boundary condition given by

$$\frac{1}{r} \frac{\partial u}{\partial \theta} + \lambda u = 0, \quad \theta = 0, \quad (4)$$

$$\frac{\partial u}{\partial \theta} = 0, \quad \theta = \alpha, \quad (5)$$

into the simpler boundary conditions given by

$$v = 0, \quad \theta = 0, \quad (6)$$

$$\frac{\partial v}{\partial \theta} = 0, \quad \theta = \alpha. \quad (7)$$

From Stoker's paper on water waves [9], we are led immediately to the following operator

$$L = \prod_{p=1}^n (\alpha_{2p} D + \lambda) \quad (8)$$

where

$$\alpha_k = e^{-i(\pi/2)(k/n+1)} \quad k = 2, 4, \dots, 2n \quad (9)$$

and D is the complex operator d/dz . D has the value $\partial/\partial x$ along the real axis, and the value $\partial/\partial(iy)$ along the imaginary axis. Hence, the appropriate operator for a wedge of internal angle $\pi/2n$ with an impedance boundary condition on the upper surface and a zero normal derivative condition on the lower surface becomes

$$L = - \prod_{p=1}^n \left\{ \sin \frac{\pi p}{n} \frac{\partial}{\partial x} + \cos \frac{\pi p}{n} \frac{\partial}{\partial y} - \lambda \right\}. \quad (10)$$

The transformation that simplifies the original boundary conditions (4) and (5) is

$$v = Lu. \quad (11)$$

We make the important observation that the same operator and transformation applies to the case of a wedge of twice the exterior angle when the lower face

has the same impedance boundary condition as the upper face. This fact follows immediately from symmetry.

III. DIFFRACTED FAR FIELD DUE TO AN INCIDENT PLANE WAVE

Consider a wedge of angle $\pi/2n$, where n is an integer, defined by the surfaces $\theta = 0$ and $\theta = \alpha$ as shown in Fig. 1.

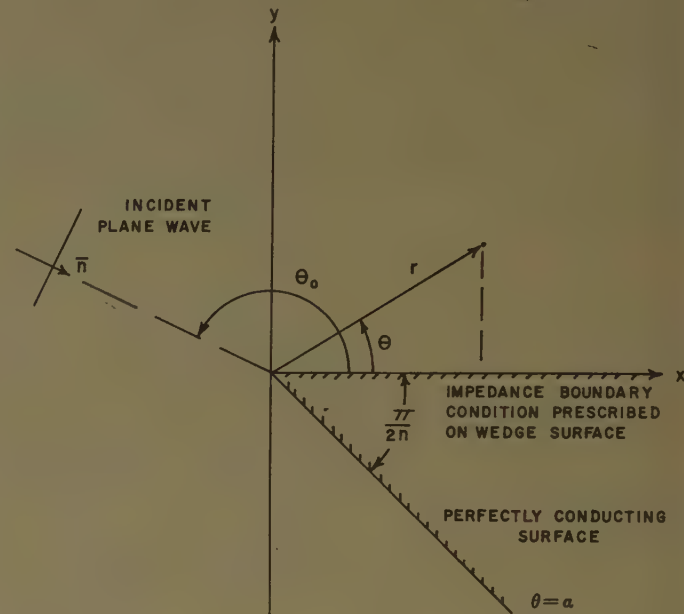


Fig. 1.

In the angular region $0 \leq \theta \leq \alpha$ we assume that we have free space. A plane wave u_I , whose magnetic vector is linearly polarized in the z -direction, is incident on the wedge. If the angle of incidence is θ_0 and the direction of the normal to the plane wave front is as shown in Fig. 1, then

$$u_I = U_0 e^{-ik(x \cos \theta_0 + y \sin \theta_0)} \quad (12)$$

where U_0 is the magnitude of the incident plane wave and the time dependence $e^{-i\omega t}$ is omitted for convenience. We assume that the boundary conditions on the wedge are given by

$$\frac{1}{r} \frac{\partial u}{\partial \theta} + \lambda u = 0, \quad \theta = 0, \quad (13)$$

$$\frac{\partial u}{\partial \theta} = 0, \quad \theta = \alpha, \quad (14)$$

where u is the z -component of the magnetic vector and λ has already been defined. It is easily shown that in two dimensions the field component u is independent of z and completely determines the electromagnetic field. The field component $H_z \equiv u$ satisfies the wave equation

$$(\nabla^2 + k^2)u = 0 \quad (15)$$

where ∇^2 is the rectangular Laplacian and k is the propagation constant of free space. In addition, u is required to be bounded as the edge is approached. From (12),

$$\tilde{B}(\theta, \theta_0) = \frac{\frac{n}{4n-1} \sin \frac{n\theta_0}{4n-1} \sin \frac{n\pi}{4n-1} \sin \frac{n\theta}{4n-1}}{\left[\cos \frac{n(\theta - \theta_0)}{4n-1} - \cos \frac{n\pi}{4n-1} \right] \left[\cos \frac{n(\theta + \theta_0)}{4n-1} - \cos \frac{n\pi}{4n-1} \right]} \quad (24)$$

(10), and (11) we obtain the form of the incident plane wave in the transformed problem. We have

$$v_I = V_0 e^{-ik(x \cos \theta_0 + y \sin \theta_0)} \quad (16)$$

where

$$V_0 = ikU_0 \prod_{p=1}^n \left\{ \sin \left(\frac{p\pi}{n} + \theta_0 \right) + \sin \theta_1 \right\} \quad (17)$$

and

$$\lambda = ik \sin \theta_1. \quad (18)$$

The quantity u in the far field can be written as the sum of two terms. The first term can be found by using geometrical optics and is composed of incident and reflected waves. The second term is due to the diffracted field. We have

$$u = (u)_{\text{geom}} + (u)_{\text{diff}} \quad (19)$$

where $(u)_{\text{geom}}$ is assumed to be known and $(u)_{\text{diff}}$ is to be found. The transformed quantity v can be written in the following form for large r

$$v = (v)_{\text{geom}} + [(v)_{\text{diff}}]_F + [(v)_{\text{diff}}]_S. \quad (20)$$

The first term can be found *exactly* by using geometrical optics and (10) and (11), and is assumed to be known. The diffracted far-field solution of the transformed problem which is finite at the edge of the wedge and is denoted by the subscript F is assumed to be known and can be obtained by standard methods. The far-field solution of the transformed problem which is singular at the edge of the wedge, and is denoted by the subscript S , is the asymptotic form of¹

$$[v]_S = \sum_{p=1}^{2n-1} c_p H_{(2p-1)n/(4n-1)}^{(1)}(kr) \sin \frac{(2p-1)n}{4n-1} \theta. \quad (21)$$

The far-field expression for the finite solution is given by

$$[(v)_{\text{diff}}]_F \approx \sqrt{\frac{2}{\pi kr}} e^{ikr - i(\pi/4)} \tilde{B}(\theta, \theta_0) \quad (22)$$

where

$$\tilde{B}(\theta, \theta_0) = kU_0 \tilde{B}(\theta, \theta_0) \prod_{p=1}^n \left\{ \sin \left(\frac{p\pi}{n} + \theta_0 \right) + \sin \theta_1 \right\} \quad (23)$$

and

¹ The series must be terminated at $p=2n-1$ in order to assure the previously required finiteness of u at the edge, as can be seen from the number of differentiations required in going from u to v .

\tilde{B} is the far-field amplitude for the boundary value problem in (6) and (7) due to a *unit* amplitude incident plane wave. The far-field expression for the singular solution is given by

$$[(v)_{\text{diff}}]_S \approx \sqrt{\frac{2}{\pi kr}} e^{ikr - i(\pi/4)} \sum_{p=1}^{2n-1} \tilde{c}_{pn} \sin \frac{n(2p-1)}{4n-1} \theta \quad (25)$$

where

$$\tilde{c}_{pn} = c_p e^{-i[n(2p-1)/(4n-1)]\pi/2}. \quad (26)$$

We expect *a priori* that the far-field representation of the original field u is of the form

$$(u)_{\text{diff}} \approx \sqrt{\frac{2}{\pi kr}} e^{ikr - i(\pi/4)} \tilde{n}(\theta) \quad (27)$$

where $\tilde{n}(\theta)$ is unknown. Consider now only the diffracted part of the far-field expressions. We have

$$(v)_{\text{diff}} = L\{(u)_{\text{diff}}\} \quad (28)$$

or

$$[(v)_{\text{diff}}]_F + [(v)_{\text{diff}}]_S = - \prod_{p=1}^n \left\{ \sin \frac{p\pi}{n} \frac{\partial}{\partial x} + \cos \frac{p\pi}{n} \frac{\partial}{\partial y} - \lambda \right\} (u)_{\text{diff}}. \quad (29)$$

But in the far field

$$\frac{\partial}{\partial x} = \cos \theta \frac{\partial}{\partial r} - \frac{1}{r} \sin \theta \frac{\partial}{\partial \theta} \approx \cos \theta \frac{\partial}{\partial r} \approx ik \cos \theta, \quad (30)$$

$$\frac{\partial}{\partial y} = \sin \theta \frac{\partial}{\partial r} + \frac{1}{r} \cos \theta \frac{\partial}{\partial \theta} \approx \sin \theta \frac{\partial}{\partial r} \approx ik \sin \theta, \quad (31)$$

and hence,

$$[(v)_{\text{diff}}]_F + [(v)_{\text{diff}}]_S = - \prod_{p=1}^n \left\{ \sin \left(\frac{p\pi}{n} + \theta \right) \frac{\partial}{\partial r} - \lambda \right\} (u)_{\text{diff}}. \quad (32)$$

If we substitute (18), (22), (25), and (27) into the above we obtain

$$\tilde{n}(\theta) = \frac{\left(\frac{i}{k} \right) \left[\tilde{B}(\theta, \theta_0) + \sum_{p=1}^{2n-1} \tilde{c}_{pn} \sin \frac{n(2p-1)}{4n-1} \theta \right]}{\prod_{p=1}^n \left[\sin \left(\frac{p\pi}{n} + \theta \right) - \sin \theta_1 \right]}. \quad (33)$$

The constants \tilde{c}_{pn} are as yet unknown.

Notice, however, that the denominator of (33) vanishes for certain values of θ . Let the roots of the equation

$$\prod_{p=1}^n \left[\sin \left(\frac{\pi p}{n} + \theta \right) - \sin \theta_1 \right] = 0 \quad (34)$$

be denoted by β_j where j is an integer. There are $2n-1$ values of j [i.e., there are $2n-1$ roots of (34) whose values are given in (64)] in the interval $0 < \theta < 2\pi - \pi/2n$ and at these roots there must be no poles of $\tilde{n}(\theta)$. The reason that the far-field amplitude $\tilde{n}(\theta)$ can have no infinities at the roots of (34) is that these would contribute

inadmissible shadow lines.² Hence the numerator of (33) must vanish at the roots β_j of the denominator. Therefore we can write

$$(u)_{\text{diff}} = \sqrt{\frac{2}{k\pi r}} e^{i(kr + \pi/4)} U_0 \frac{\prod_{p=1}^n \left\{ \sin \left(\frac{p\pi}{n} + \theta_0 \right) + \sin \theta_1 \right\} N_{2n-1}}{\prod_{p=1}^n \left\{ \sin \left(\frac{p\pi}{n} + \theta \right) - \sin \theta_1 \right\} \Delta_{2n-1}} \quad (35)$$

where

$$N_{2n-1} = \begin{vmatrix} \tilde{B}(\theta, \theta_0) & \sin \frac{n}{4n-1} \theta & \sin 3 \frac{n}{4n-1} \theta & \cdots & \sin (4n-3) \frac{n}{4n-1} \theta \\ \tilde{B}(\beta_1, \theta_0) & \sin \frac{n}{4n-1} \beta_1 & \sin 3 \frac{n}{4n-1} \beta_1 & \cdots & \sin (4n-3) \frac{n}{4n-1} \beta_1 \\ \tilde{B}(\beta_2, \theta_0) & \sin \frac{n}{4n-1} \beta_2 & \sin 3 \frac{n}{4n-1} \beta_2 & \cdots & \sin (4n-3) \frac{n}{4n-1} \beta_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{B}(\beta_{2n-1}, \theta_0) & \sin \frac{n}{4n-1} \beta_{2n-1} & \sin 3 \frac{n}{4n-1} \beta_{2n-1} & \cdots & \sin (4n-3) \frac{n}{4n-1} \beta_{2n-1} \end{vmatrix}, \quad (36)$$

$$\Delta_{2n-1} = \begin{vmatrix} \sin \frac{n}{4n-1} \beta_1 & \sin 3 \frac{n}{4n-1} \beta_1 & \cdots & \sin (4n-3) \frac{n}{4n-1} \beta_1 \\ \sin \frac{n}{4n-1} \beta_2 & \sin 3 \frac{n}{4n-1} \beta_2 & \cdots & \sin (4n-3) \frac{n}{4n-1} \beta_2 \\ \vdots & \vdots & \ddots & \vdots \\ \sin \frac{n}{4n-1} \beta_{2n-1} & \sin 3 \frac{n}{4n-1} \beta_{2n-1} & \cdots & \sin (4n-3) \frac{n}{4n-1} \beta_{2n-1} \end{vmatrix}. \quad (37)$$

The roots β_j are here regarded as being enumerated in some arbitrary but fixed order. We first consider the determinant Δ_{2n-1} . It can be shown that its value is given by

$$\begin{aligned} \Delta_{2n-1} = & (2)^{2(2n-1)(n-1)} \sin x_1 \sin x_2 \sin x_3 \cdots \sin x_{2n-1} \\ & \cdot [\cos^2 x_2 - \cos^2 x_1] [\cos^2 x_3 - \cos^2 x_1] [\cos^2 x_4 - \cos^2 x_1] \cdots [\cos^2 x_{2n-1} - \cos^2 x_1] \\ & \quad [\cos^2 x_3 - \cos^2 x_2] [\cos^2 x_4 - \cos^2 x_2] \cdots [\cos^2 x_{2n-1} - \cos^2 x_2] \\ & \quad [\cos^2 x_4 - \cos^2 x_3] \cdots [\cos^2 x_{2n-1} - \cos^2 x_3] \\ & \quad \vdots \\ & \quad [\cos^2 x_{2n-1} - \cos^2 x_{2n-2}] \end{aligned} \quad (38)$$

² In view of the previously noted restriction (s real, $0 < s < 1$), only real roots occur in (34). For more general values of s , complex roots can occur (as for example, for the case of surface waves). A similar analysis can still be applied, though in more complicated form.

where

$$x_j = \frac{n}{4n-1} \beta_j. \quad (39)$$

A principal minor of the numerator is of the form

$$(-1)^j \widetilde{B}(\beta_j, \theta_0) \left| \begin{array}{cccc} \sin \phi & \sin 3\phi & \cdots & \sin (4n-3)\phi \\ \sin x_1 & \sin 3x_1 & & \sin (4n-3)x_1 \\ \sin x_2 & \sin 3x_2 & & \sin (4n-3)x_2 \\ \vdots & \vdots & & \vdots \\ \sin x_{j-1} & \sin 3x_{j-1} & & \sin (4n-3)x_{j-1} \\ \sin x_{j+1} & \sin 3x_{j+1} & & \sin (4n-3)x_{j+1} \\ \vdots & \vdots & & \vdots \\ \sin x_{2n-1} & \sin 3x_{2n-1} & \cdots & \sin (4n-3)x_{2n-1} \end{array} \right| \quad (40)$$

where

$$\phi = \frac{n}{4n-1} \theta. \quad (41)$$

Note that the j th row is omitted. Expanding the above as before we have

[illegible]

Dividing numerator by denominator, and cancelling all common terms, we obtain

$$(u)_{\text{diff}} = \sqrt{\frac{2}{\pi k r}} e^{i(kr + (\pi/4))} U_0 \frac{n}{4n-1} \sin \frac{n\theta}{4n-1} \sin \frac{n\theta_0}{4n-1} \sin \frac{n\pi}{4n-1} \frac{\prod_{p=1}^n \left\{ \sin \left(\frac{p\pi}{n} + \theta_0 \right) + \sin \theta_1 \right\}}{\prod_{p=1}^n \left\{ \sin \left(\frac{p\pi}{n} + \theta \right) - \sin \theta_1 \right\}} \\ \cdot \left[+ \frac{1}{\left[\cos \frac{n(\theta - \theta_0)}{4n-1} - \cos \frac{n\pi}{4n-1} \right] \left[\cos \frac{n(\theta + \theta_0)}{4n-1} - \cos \frac{n\pi}{4n-1} \right]} \right. \\ - \frac{1}{\left[\cos \frac{n(\beta_1 - \theta_0)}{4n-1} - \cos \frac{n\pi}{4n-1} \right] \left[\cos \frac{n(\beta_1 + \theta_0)}{4n-1} - \cos \frac{n\pi}{4n-1} \right]} \frac{[\cos^2 x_2 - \cos^2 \phi][\cos^2 x_3 - \cos^2 \phi] \cdots [\cos^2 x_{2n-1} - \cos^2 \phi]}{[\cos^2 x_2 - \cos^2 x_1][\cos^2 x_3 - \cos^2 x_1] \cdots [\cos^2 x_{2n-1} - \cos^2 x_1]} \\ - \frac{1}{\left[\cos \frac{n(\beta_2 - \theta_0)}{4n-1} - \cos \frac{n\pi}{4n-1} \right] \left[\cos \frac{n(\beta_2 + \theta_0)}{4n-1} - \cos \frac{n\pi}{4n-1} \right]} \frac{[\cos^2 x_1 - \cos^2 \phi][\cos^2 x_3 - \cos^2 \phi] \cdots [\cos^2 x_{2n-1} - \cos^2 \phi]}{[\cos^2 x_1 - \cos^2 x_2][\cos^2 x_3 - \cos^2 x_2] \cdots [\cos^2 x_{2n-1} - \cos^2 x_2]} \\ - \frac{1}{\left[\cos \frac{n(\beta_3 - \theta_0)}{4n-1} - \cos \frac{n\pi}{4n-1} \right] \left[\cos \frac{n(\beta_3 + \theta_0)}{4n-1} - \cos \frac{n\pi}{4n-1} \right]} \frac{[\cos^2 x_1 - \cos^2 \phi][\cos^2 x_2 - \cos^2 \phi] \cdots [\cos^2 x_{2n-1} - \cos^2 \phi]}{[\cos^2 x_1 - \cos^2 x_3][\cos^2 x_2 - \cos^2 x_3] \cdots [\cos^2 x_{2n-1} - \cos^2 x_3]} \\ \left. - \cdots \right]. \quad (43)$$

We can write (43) in a different form involving only sines. Further simplification yields the following compact expression:

$$(u)_{\text{diff}} = \sqrt{\frac{2}{\pi k r}} e^{i(kr + \pi/4)} U_0 \frac{n}{4(4n-1)} \sin \frac{n\theta}{4n-1} \sin \frac{n\theta_0}{4n-1} \sin \frac{n\pi}{4n-1} \frac{\prod_{p=1}^n \left\{ \sin \left(\frac{p\pi}{n} + \theta_0 \right) + \sin \theta_1 \right\}}{\prod_{p=1}^n \left\{ \sin \left(\frac{p\pi}{n} + \theta \right) - \sin \theta_1 \right\}} \left[+ \frac{1}{\sin \frac{n(\theta - \theta_0 + \pi)}{2(4n-1)} \sin \frac{n(\theta - \theta_0 - \pi)}{2(4n-1)} \sin \frac{n(\theta + \theta_0 + \pi)}{2(4n-1)} \sin \frac{n(\theta + \theta_0 - \pi)}{2(4n-1)}} - \sum_{j=1}^{2n-1} \frac{1}{\sin \frac{n(\beta_j - \theta_0 + \pi)}{2(4n-1)} \sin \frac{n(\beta_j - \theta_0 - \pi)}{2(4n-1)} \sin \frac{n(\beta_j + \theta_0 + \pi)}{2(4n-1)} \sin \frac{n(\beta_j + \theta_0 - \pi)}{2(4n-1)}} \frac{\prod_{p=1}^{2n-1} \sin \frac{n(\beta_p - \theta)}{4n-1} \sin \frac{n(\beta_p + \theta)}{4n-1}}{\prod_{p=1}^{2n-1} \sin \frac{n(\beta_p - \beta_j)}{4n-1} \sin \frac{n(\beta_p + \beta_j)}{4n-1}} \right] \quad (44)$$

This result is the diffracted far field due to an incident plane wave for a wedge of internal angle $\pi/2n$ (i.e., in an external angular space $2\pi - \pi/2n$) when an impedance boundary condition is prescribed on the upper face and a Neumann condition is prescribed on the lower. It is also the result for the diffracted far field for a wedge in an external space $4\pi - \pi/n$ ($0 < \theta < 4\pi - \pi/n$) when the same impedance boundary condition is prescribed on both surfaces, provided the incident wave arrives along the line of symmetry of the wedge.

IV. DIFFRACTED FAR FIELD DUE TO A LINE SOURCE LOCATED AT THE TIP

Consider now a wedge of angle $\pi/2n$, where n is an integer, defined by the surfaces $\theta=0$ and $\theta=\alpha$ as shown in Fig. 2. In the angular region $0 \leq \theta \leq \alpha$ we assume that we have free space. A magnetic line source is located at the tip of the wedge and has coordinates $x=0, y=0$ (see Fig. 2). We assume that the boundary conditions on the wedge are given by

$$\frac{1}{r} \frac{\partial u}{\partial \theta} + \lambda u = 0, \quad \theta = 0, \quad (45)$$

$$\frac{\partial u}{\partial \theta} = 0, \quad \theta = \alpha, \quad (46)$$

where u is the z -component of the magnetic vector and λ has already been defined. The field component $H_z \equiv u$ satisfies the wave equation

$$(\nabla^2 + k^2)u = -4\pi\delta(x-0)\delta(y-0) \quad (47)$$

where ∇^2 is the rectangular Laplacian, k is the propaga-

tion constant of free space, and δ is the Dirac delta function.

The radiated far field is found in a manner similar to that used for the incident wave plane. Again we first introduce the operator

$$v = Lu \quad (48)$$

where

$$L = - \prod_{p=1}^n \left\{ \sin \frac{\pi p}{n} \frac{\partial}{\partial x} + \cos \frac{\pi p}{n} \frac{\partial}{\partial y} - \lambda \right\}. \quad (49)$$

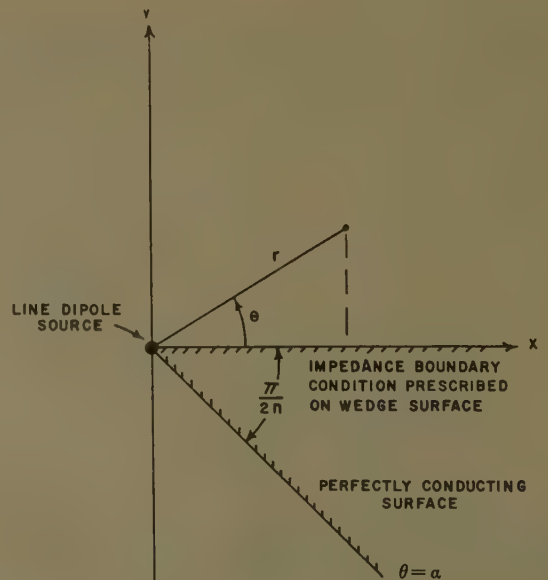


Fig. 2.

The inverse operation $u = L^{-1}v$ involves n integrations. Let v_s be that part of v which arises from performing integration on the source term. Therefore, v_s must be of the form

$$v_s = c_n H_n^{(1)}(kr) \sin n\theta \quad (50)$$

since the solution of (47) is given by the Green's function

$$u_s = i\pi H_0^{(1)}(kr). \quad (51)$$

In the above equations $H_n^{(1)}(kr)$ is a Hankel function of the first kind of order n and the subscript S indicates that these are source terms; c_n is a normalization constant. The solution of the transformed problem is given by¹

$$v = c_n H_n^{(1)}(kr) \sin n\theta + \sum_{p=1}^{2n-1} c_p H_{(2p-1)n/(4n-1)}^{(1)}(kr) \sin \frac{(2p-1)n}{4n-1} \theta \quad (52)$$

and has the far-field representation

$$(v)_{ff} \approx c_n \sqrt{\frac{2}{\pi kr}} e^{i(kr-\pi/4)} e^{-i(n\pi/2)} \sin n\theta + \sqrt{\frac{2}{\pi kr}} e^{i(kr-\pi/4)} \sum_{p=1}^{2n-1} c_p \cdot \exp\left(-i \frac{n(2p-1)}{4n-1} \frac{\pi}{2}\right) \sin \frac{n(2p-1)}{4n-1} \theta, \quad (53)$$

where v_{ff} denotes the far-field form of v . We expect *a priori* that the far-field form of the original function u is

$$(u)_{ff} \approx \sqrt{\frac{2}{\pi kr}} e^{i(kr-\pi/4)} \tilde{n}(\theta) \quad (54)$$

where $\tilde{n}(\theta)$ is unknown. If we now consider the far-field form of v we have

$$(v)_{ff} = - \prod_{p=1}^n \left\{ \sin\left(\frac{\pi p}{n} + \theta\right) \frac{\partial}{\partial r} - \lambda \right\} (u)_{ff}. \quad (55)$$

Substituting from (54) and (55) we obtain

$$c_n e^{-i(n\pi/2)} \sin n\theta + \sum_{p=1}^{2n-1} c_p \cdot \exp\left(-i \frac{n(2p-1)}{4n-1} \frac{\pi}{2}\right) \sin \frac{n(2p-1)}{4n-1} \theta = -ik \prod_{p=1}^n \left\{ \sin\left(\frac{\pi p}{n} + \theta\right) - \sin \theta_1 \right\} \tilde{n}(\theta) \quad (56)$$

where we have set

$$\lambda = ik \sin \theta_1. \quad (57)$$

Hence

$$(u)_{ff} = \sqrt{\frac{2}{\pi kr}} e^{i(kr+\pi/4)} \left(\frac{1}{k}\right) \frac{\left[\tilde{c}_n \sin n\theta + \sum_{p=1}^{2n-1} \tilde{c}_{2p} \sin \frac{n(2p-1)}{4n-1} \theta \right]}{\prod_{p=1}^n \left[\sin\left(\frac{\pi p}{n} + \theta\right) - \sin \theta_1 \right]} \quad (58)$$

where

$$\tilde{c}_n = c_n e^{-i(n\pi/2)} \quad (59)$$

$$\tilde{c}_{2p} = c_p \exp\left(-i \frac{n(2p-1)}{4n-1} \frac{\pi}{2}\right). \quad (60)$$

Using the same argument as in Section III, we note that $(u)_{ff}$ can be written in the form

$$(u)_{ff} = \sqrt{\frac{2}{\pi kr}} e^{i(kr+\pi/4)} \left(\frac{\tilde{c}_n}{k}\right) \frac{N_{2n-1}}{\Delta_{2n-1}} \prod_{p=1}^n \left[\sin\left(\frac{\pi p}{n} + \theta\right) - \sin \theta_1 \right] \quad (61)$$

where

$$N_{2n-1} = \begin{vmatrix} \sin n\theta & \sin \frac{n\theta}{4n-1} & \sin 3 \frac{n\theta}{4n-1} & \cdots & \sin (4n-3) \frac{n\theta}{4n-1} \\ \sin n\beta_1 & \sin \frac{n\beta_1}{4n-1} & \sin 3 \frac{n\beta_1}{4n-1} & \cdots & \sin (4n-3) \frac{n\beta_1}{4n-1} \\ \sin n\beta_2 & \sin \frac{n\beta_2}{4n-1} & \sin 3 \frac{n\beta_2}{4n-1} & \cdots & \sin (4n-3) \frac{n\beta_2}{4n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sin n\beta_{2n-1} & \sin \frac{n\beta_{2n-1}}{4n-1} & \sin 3 \frac{n\beta_{2n-1}}{4n-1} & \cdots & \sin (4n-3) \frac{n\beta_{2n-1}}{4n-1} \end{vmatrix}, \quad (62)$$

$$\Delta_{2n-1} = \begin{vmatrix} \sin \frac{n\beta_1}{4n-1} & \sin 3 \frac{n\beta_1}{4n-1} & \cdots & \sin (4n-3) \frac{n\beta_1}{4n-1} \\ \sin \frac{n\beta_2}{4n-1} & \sin 3 \frac{n\beta_2}{4n-1} & \cdots & \sin (4n-3) \frac{n\beta_2}{4n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sin \frac{n\beta_{2n-1}}{4n-1} & \sin 3 \frac{n\beta_{2n-1}}{4n-1} & \cdots & \sin (4n-3) \frac{n\beta_{2n-1}}{4n-1} \end{vmatrix}. \quad (63)$$

In the above determinants the β_j 's are the $(2n-1)$ roots of (34) which lead to inadmissible poles and these are given by

$$\beta_j = \theta_1 + \pi + \frac{\pi}{n}(n-j), \quad j = 1, \cdots, n,$$

$$\beta_{n+s} = -\theta_1 + \frac{\pi}{n}(n-s), \quad s = 1, \cdots, n-1. \quad (64)$$

By using a procedure similar to that in Section III we have at once that

$$(u)_{ff} = \sqrt{\frac{2}{\pi k r}} e^{i(kr+\pi/4)} \frac{\left(\frac{\tilde{c}_n}{k}\right) \sin \frac{n\theta}{4n-1}}{\prod_{p=1}^n \left[\sin \left(\frac{\pi p}{n} + \theta \right) - \sin \theta_1 \right]} \cdot \left[+ \frac{\sin n\theta}{\sin \frac{n\theta}{4n-1}} - \sum_{j=1}^{2n-1} \frac{\sin n\beta_j}{\sin \frac{n\beta_j}{4n-1}} \right. \\ \left. \cdot \frac{\prod_{\substack{p=1 \\ (p \neq j)}}^{2n-1} \sin \frac{n(\beta_p - \theta)}{4n-1} \sin \frac{n(\beta_p + \theta)}{4n-1}}{\prod_{p=1}^{2n-1} \sin \frac{n(\beta_p - \beta_j)}{4n-1} \sin \frac{n(\beta_p + \beta_j)}{4n-1}} \right] \quad (65)$$

where

$$\tilde{c}_n = -\pi k \frac{1}{2^{n-1}}. \quad (66)$$

³ The calculation involves noting that

$$i\pi H_0^{(1)}(kr) \rightarrow \sqrt{\frac{2}{\pi k r}} e^{i(kr+\pi/4)}$$

and

$$\prod_{p=1}^n \sin \left(\frac{\pi p}{n} + \theta \right) = -\frac{1}{2^{n-1}} \sin n\theta.$$

The above constant \tilde{c}_n can be found by considering the limiting case of $\theta_1=0$ ($\lambda=0$) where the solution is $i\pi H_0^{(1)}(kr)$ independently of n .³ This result is the diffracted far field due to a line source located at the tip of a wedge of angle $\pi/2n$ (exterior angle $2\pi - \pi/2n$) when an impedance condition is prescribed on the upper face and a Neumann boundary condition is prescribed on the lower face. (It is also the result for the diffracted far field for a wedge of exterior angle $(4\pi - \pi/n)$ when the same impedance boundary condition is prescribed on both surfaces.)

V. APPLICATIONS

A. Right-angled Wedge Excited by a Line Source Located at the Tip with One Absorbing Face and One Perfectly Conducting Face.

For a right-angled wedge $n=1$. The appropriate root is $\beta_1 = \theta_1 + \pi$ where $0 \leq \theta_1 \leq \pi$. Using (65) we find that the desired radiated far field is given by

$$(u)_{ff} = \sqrt{\frac{2}{\pi k r}} e^{i(kr+\pi/4)} \frac{\left(\frac{\tilde{c}_1}{k}\right) \sin \frac{\theta}{3}}{[\sin(\pi + \theta) - \sin \theta_1]} \cdot \left[\frac{\sin \theta}{\sin \frac{\theta}{3}} - \frac{\sin \beta_1}{\sin \frac{\beta_1}{3}} \right] \quad (67)$$

where $\tilde{c}_1 = -\pi k$.

B. Forty-Five-Degree Wedge Excited by a Line Source Located at the Tip with One Absorbing Face and One Perfectly Conducting Face.

For a forty-five-degree wedge, $n=2$. For an absorber, the appropriate roots are $\beta_1 = \theta_1 + \pi$, $\beta_2 = \pi/2 - \theta_1$ and $\beta_3 = 3\pi/2 + \theta_1$. Using (65) we find that the desired radiated far field is given by

$$(u)_{ff} = \sqrt{\frac{2}{\pi k r}} e^{i(kr + \pi/4)} \frac{\left(\frac{\tilde{c}_2}{k}\right) \sin \frac{2\theta}{7}}{\left[\sin\left(\frac{\pi}{2} + \theta\right) - \sin \theta_1\right] [\sin(\pi + \theta) - \sin \theta_1]} \cdot \left[\frac{\sin 2\theta}{\sin \frac{2\theta}{7}} - \frac{\sin 2\beta_1}{\sin \frac{2\beta_1}{7}} \frac{\sin \frac{2}{7}(\beta_2 - \theta) \sin \frac{2}{7}(\beta_2 + \theta) \sin \frac{2}{7}(\beta_3 - \theta) \sin \frac{2}{7}(\beta_3 + \theta)}{\sin \frac{2}{7}(\beta_2 - \beta_1) \sin \frac{2}{7}(\beta_2 + \beta_1) \sin \frac{2}{7}(\beta_3 - \beta_1) \sin \frac{2}{7}(\beta_3 + \beta_1)} \right. \\ \left. - \frac{\sin 2\beta_2}{\sin \frac{2\beta_2}{7}} \frac{\sin \frac{2}{7}(\beta_1 - \theta) \sin \frac{2}{7}(\beta_1 + \theta) \sin \frac{2}{7}(\beta_3 - \theta) \sin \frac{2}{7}(\beta_3 + \theta)}{\sin \frac{2}{7}(\beta_1 - \beta_2) \sin \frac{2}{7}(\beta_1 + \beta_2) \sin \frac{2}{7}(\beta_3 - \beta_2) \sin \frac{2}{7}(\beta_3 + \beta_2)} \right. \\ \left. - \frac{\sin 2\beta_3}{\sin \frac{2\beta_3}{7}} \frac{\sin \frac{2}{7}(\beta_1 - \theta) \sin \frac{2}{7}(\beta_1 + \theta) \sin \frac{2}{7}(\beta_2 - \theta) \sin \frac{2}{7}(\beta_2 + \theta)}{\sin \frac{2}{7}(\beta_1 - \beta_3) \sin \frac{2}{7}(\beta_1 + \beta_3) \sin \frac{2}{7}(\beta_2 - \beta_3) \sin \frac{2}{7}(\beta_2 + \beta_3)} \right] \quad (68)$$

where

$$\tilde{c}_2 = -(k/2)\pi. \quad (69)$$

VI. APPLICATION OF METHOD TO SIMPLE BOUNDARY CONDITIONS AS IN THE CASE OF PERFECT CONDUCTORS

The method employed above can also be employed with modification in treating the more ordinary classical diffraction problem, arising in the case when both wedge faces are perfectly conducting. Here we illustrate the method in the cases of a half plane and of a right-angled wedge.

A. Diffraction by a Perfectly Conducting Half Plane

Consider the problem described below:

$$u_{xx} + u_{yy} + k^2 u = 0, \quad 0 \leq \theta \leq 2\pi. \quad (70)$$

$$u = e^{ik(x \cos \theta_0 + y \sin \theta_0)} + \text{outgoing wave}. \quad (71a)$$

$$u(-\infty, y) = e^{ik(x \cos \theta_0 + y \sin \theta_0)}, \quad -\frac{\pi}{2} < \theta_0 < +\frac{\pi}{2}. \quad (71b)$$

$$u(x, 0) = 0, \quad x > 0. \quad (72)$$

$$u \text{ finite at } x = 0, y = 0. \quad (73)$$

Here⁴ we let $-\pi < \theta_0 < \pi$.

Let

$$v = \frac{\partial u}{\partial x} - ik \cos \theta_0 u. \quad (74)$$

Then v is an outgoing wave function, which obeys the

radiation condition. Also, from (72) and (74) we have that $v(x, 0) = 0$, $x > 0$. Therefore (72) implies that the most general form of v is given by

$$v = \sum_{n=1}^{\infty} a_n H_{n/2}^{(1)}(kr) \sin \frac{n}{2} \theta \quad (75a)$$

where the a_n are to be determined. Now (74) shows that $v = 0(u/x)$ near the origin; therefore (73) and (75a) imply that $u \rightarrow 0$ at origin, and also that $a_n = 0$, $n > 1$. Therefore

$$v = a_1(\theta_0) H_{1/2}^{(1)}(kr) \sin \frac{\theta}{2} = \frac{\partial u}{\partial x} - ik \cos \theta_0 u. \quad (75b)$$

We have yet to determine a_1 .

Now let

$$u = u_{\text{geom}} + u^{(1)} \quad (76)$$

and assume that

$$u^{(1)} \rightarrow \tilde{H}_0^{(1)}(kr) u(\theta), \quad r \rightarrow \infty \quad (77)$$

where

$$\tilde{H}_0^{(1)}(kr) \equiv \sqrt{\frac{2}{\pi k r}} e^{ikr - i(\pi/4)}. \quad (78)$$

Eq. (77) holds for almost all fixed θ . We also know that (74) holds, and that

$$v \rightarrow \tilde{H}_0^{(1)}(kr) e^{-i(\pi/4)} a_1 \sin \frac{\theta}{2} + O\left(\frac{1}{\sqrt{r}}\right). \quad (79)$$

But

$$\frac{\partial}{\partial x} = \cos \theta \frac{\partial}{\partial r} - \frac{1}{r} \sin \theta \frac{\partial}{\partial \theta}, \quad (80)$$

⁴ The definition of the angle θ_0 employed in this section differs from that used in earlier sections. Henceforth θ_0 means the angle between the direction of travel of the plane wave and the positive x -axis

while for large r ,

$$\begin{aligned}\frac{\partial}{\partial r} \tilde{H}_0^{(1)}(kr) &\sim ik \tilde{H}_0^{(1)}(kr) + O\left(\frac{1}{\sqrt{r}}\right), \\ \frac{1}{r} \tilde{H}_0^{(1)}(kr) &\sim O\left(\frac{1}{r^{3/2}}\right).\end{aligned}\quad (81)$$

Therefore, for functions of the type illustrated in (77) and (79), $\partial/\partial x \rightarrow ik \cos \theta$. Using (74) and (79) the coefficient of $H_0^{(1)}$ in (79) now becomes

$$a_1 e^{-i(\pi/4)} \sin \frac{\theta}{2} = ik(\cos \theta - \cos \theta_0) u(\theta).$$

Hence,

$$u(\theta) = \frac{a_1 \sin \frac{\theta}{2} e^{-i(\pi/4)}}{ik(\cos \theta - \cos \theta_0)} \quad (82)$$

where, according to (77), $u(\theta)$ is the (physically interesting) complex far-field amplitude of the diffracted wave. The θ dependence is quite explicit. We need the constant a_1 , which may depend only on θ_0 . To emphasize the dependence on the angle of incidence we rewrite (82) as follows:

$$u(\theta, \theta_0) \equiv u(\theta) = \frac{a_1(\theta_0) \sin \frac{\theta}{2} e^{-i(\pi/4)}}{ik(\cos \theta - \cos \theta_0)}. \quad (83)$$

Now by reciprocity

$$u(\theta, \theta_0) = u(\theta_0 + \pi, \theta - \pi). \quad (84)$$

Here we have employed the convention that angles of incidence lie in the range $(-\pi, \pi)$. Applying (84) to (83) we find that

$$\begin{aligned}\frac{a_1(\theta_0) \sin \frac{\theta}{2}}{ik(\cos \theta - \cos \theta_0)} &= \frac{a_1(\theta - \pi) \sin \frac{\theta_0 + \pi}{2}}{ik[\cos(\theta_0 + \pi) - \cos(\theta - \pi)]} \\ &= \frac{a_1(\theta - \pi) \sin \frac{\theta_0 + \pi}{2}}{ik(\cos \theta - \cos \theta_0)}\end{aligned}\quad (85)$$

i.e.,

$$\frac{a_1(\theta_0)}{\sin \frac{\theta_0 + \pi}{2}} = \frac{a_1(\theta - \pi)}{\sin \frac{\theta}{2}} = \text{constant} \quad (86)$$

or

$$a_1(\theta_0) = a_1(0) \cos \frac{\theta_0}{2} \quad (87)$$

$$u(\theta, \theta_0) = (-ik e^{-i(\pi/4)}) \frac{\cos \frac{\theta_0}{2} \sin \frac{\theta}{2} e^{-i(\pi/4)}}{ik(\cos \theta - \cos \theta_0)}. \quad (88)$$

Here we have used (87) with a value of $a_1(0)$ obtained in the next paragraph.

In order to determine $a_1(0)$ we let $\theta_0 = 0$, and employ (75b) and (74). We integrate (74), subject to (71b), and find that

$$u = e^{ikr} + a_1(0) e^{ikx} \int_{-\infty}^x e^{-ik\xi} \chi(\xi, y) d\xi \quad (89)$$

where

$$\chi(x, y) = H_{1/2}^{(1)}(kr) \sin \frac{\theta}{2} \rightarrow -i \sqrt{\frac{2}{\pi kr}} e^{ikr} \sin \frac{\theta}{2}, \quad r \rightarrow \infty. \quad (90a)$$

We have, therefore,

$$\begin{aligned}\chi(x, 0) &= 0, \quad x > 0, \\ \chi(x, 0) &= -i \sqrt{\frac{2}{\pi k}} \frac{e^{+ik|x|}}{\sqrt{|x|}}, \quad x < 0.\end{aligned}\quad (90b)$$

Imposing the boundary condition (72), and using (90b) we get for $x \rightarrow \infty$,

$$0 = e^{ikx} + a_1(0) e^{ikx} \int_{|\xi|=\infty}^0 (-i) \sqrt{\frac{2}{\pi}} \frac{e^{2ik|\xi|}}{\sqrt{k|\xi|}} \frac{d(-k|\xi|)}{k}. \quad (91)$$

Hence, finally,

$$a_1(0) = \left[\frac{i}{k\sqrt{\pi}} \int_0^\infty \frac{e^{it}}{\sqrt{t}} dt \right]^{-1} = -ik e^{-i(\pi/4)} \quad (92a)$$

so that

$$u(\theta, \theta_0) = \frac{i \cos \frac{\theta_0}{2} \sin \frac{\theta}{2}}{(\cos \theta - \cos \theta_0)}. \quad (92b)$$

The result is that for diffraction by a half plane under the condition (72), the field for large r is given by

$$u \rightarrow u_{\text{geom}} + \sqrt{\frac{2}{\pi kr}} e^{-i(\pi/4)} e^{ikr} \frac{i \cos \frac{\theta_0}{2} \sin \frac{\theta}{2}}{(\cos \theta - \cos \theta_0)}. \quad (93)$$

Similar analysis shows that when $\partial u / \partial y = 0$ is required, the screen has the form

$$u \rightarrow u_{\text{geom}} + \sqrt{\frac{2}{\pi kr}} e^{-i(\pi/4)} e^{ikr} (\text{constant}) \frac{\sin \frac{\theta_0}{2} \cos \frac{\theta}{2}}{ik(\cos \theta - \cos \theta_0)}. \quad (94)$$

The analysis given above through equation (75b) can be traced back to Lamb [7], in the case of normal incidence ($\theta_0 = \pi/2$), although Lamb omits (75a) and merely postulates (75b). No explicit formulation of the

edge condition was available to him. After obtaining (75b) Lamb expresses it in the form (again for $\theta_0 = \pi/2$)

$$\cos \theta \frac{\partial u}{\partial r} - \frac{1}{r} \sin \theta \frac{\partial u}{\partial \theta} = (\text{constant}) \frac{e^{ikr}}{\sqrt{r}} \sin \frac{\theta}{2}.$$

However, Lamb integrates the first order partial differential equation to obtain the usual Fresnel integral expression for u . The form of the resulting solution is then generalized to be valid for arbitrary θ_0 .

B. Diffraction by a Perfectly Conducting Wedge

The present procedure can also be employed in the case of a perfectly conducting wedge.

For example, consider the problem of diffraction by a perfectly conducting right-angled wedge.

$$u_{xx} + u_{yy} + k^2 u = 0. \quad (95)$$

$$u = e^{ik(x \cos \theta_0 + y \sin \theta_0)} + \text{outgoing waves}. \quad (96a)$$

$$u = u_{\text{geom}} \text{ at infinity}. \quad (96b)$$

$$u(x, 0) = 0, \quad x > 0. \quad (97a)$$

$$u(0, y) = 0, \quad y < 0. \quad (97b)$$

$$u \text{ finite at } x = 0, y = 0. \quad (98a)$$

Let

$$v = \left(\frac{\partial^2}{\partial x^2} + k^2 \cos^2 \theta_0 \right) u = - \left(\frac{\partial^2}{\partial y^2} + k^2 \sin^2 \theta_0 \right) u. \quad (98b)$$

The two expressions are equivalent by (95). Then the geometric or plane wave part of v is zero. Again, $v = 0$ on the faces of the wedge, and v must not be too singular to prevent u from being finite at the edge. Hence

$$v = a_1(\theta_0) H_{2/3}^{(1)}(kr) \sin \frac{2\theta}{3} + a_2(\theta_0) H_{4/3}^{(1)}(kr) \sin \frac{4\theta}{3}, \quad 0 < \theta < \frac{3\pi}{2} \quad (99)$$

so that for $r \rightarrow \infty$,

$$v \rightarrow \tilde{H}_0^{(1)}(kr) \left[e^{-i(\pi/3)} a_1(\theta_0) \sin \frac{2\theta}{3} + e^{-i(2\pi/3)} a_2(\theta_0) \sin \frac{4\theta}{3} \right]. \quad (100)$$

Therefore,

$$u(\theta, \theta_0) = \frac{e^{-i(\pi/3)} a_1(\theta_0) \sin \frac{2\theta}{3} + e^{-i(2\pi/3)} a_2(\theta_0) \sin \frac{4\theta}{3}}{-k^2(\cos^2 \theta - \cos^2 \theta_0)}. \quad (101)$$

Now let

$$a_1 e^{-i(\pi/3)} = \alpha_1, \quad a_2 e^{-i(2\pi/3)} = \alpha_2. \quad (102)$$

Hence

$$\begin{aligned} & \alpha_1(\theta_0) \sin \frac{2\theta}{3} + \alpha_2(\theta_0) \sin \frac{4\theta}{3} \\ &= - \left[\alpha_1(\theta - \pi) \sin \frac{2}{3}(\theta_0 + \pi) + \alpha_2(\theta - \pi) \sin \frac{4}{3}(\theta_0 + \pi) \right] \end{aligned} \quad (103)$$

since $u(\theta, \theta_0) = u(\theta_0 + \pi, \theta - \pi)$ because of reciprocity.

Now let $\theta = \theta_0 + \pi$, and deduce

$$\alpha_1(\theta_0) \sin \frac{2}{3}(\theta_0 + \pi) + \alpha_2(\theta_0) \sin \frac{4}{3}(\theta_0 + \pi) = 0$$

or

$$\alpha_2(\theta_0) = \frac{-\alpha_1(\theta_0)}{2 \cos \frac{2}{3}(\theta_0 + \pi)}. \quad (104)$$

Thus, substituting in (103),

$$\begin{aligned} & \sin \frac{2\theta}{3} \alpha_1(\theta_0) \left[1 - \frac{\cos \frac{2\theta}{3}}{\cos \frac{2}{3}(\theta_0 + \pi)} \right] \\ &= -\alpha_1(\theta - \pi) \sin \frac{2}{3}(\theta_0 + \pi) \left[1 - \frac{\cos \frac{2}{3}(\theta_0 + \pi)}{\cos \frac{2}{3}\theta} \right]. \end{aligned} \quad (105)$$

Letting $\theta = \pi$, one obtains

$$\begin{aligned} \alpha_1(\theta_0) &= \frac{-\alpha_1(0) \sin \frac{2}{3}(\theta_0 + \pi) [1 + 2 \cos \frac{2}{3}(\theta_0 + \pi)]}{\sin \frac{2\pi}{3} \left[1 + \frac{1}{2 \cos \frac{2}{3}(\theta_0 + \pi)} \right]} \\ &= \frac{-2\alpha_1(0)}{\sqrt{3}} \sin \frac{4}{3}(\theta_0 + \pi). \end{aligned} \quad (106)$$

Hence

$$\begin{aligned} u(\theta, \theta_0) &= \frac{-\frac{2}{\sqrt{3}} \alpha_1(0) \sin \frac{4}{3}(\theta_0 + \pi)}{k^2(\cos^2 \theta_0 - \cos^2 \theta)} \\ &\cdot \left[\sin \frac{2\theta}{3} - \frac{2 \sin \frac{2\theta}{3} \cos \frac{2\theta}{3}}{2 \cos \frac{2}{3}(\theta_0 + \pi)} \right] \end{aligned} \quad (107)$$

or

$$u(\theta, \theta_0) = -\frac{4\alpha_1(0)}{\sqrt{3}} \frac{\sin \frac{2}{3}(\theta_0 + \pi) \sin \frac{2}{3}\theta [\cos \frac{2}{3}(\theta_0 + \pi) - \cos \frac{2}{3}\theta]}{k^2(\cos^2 \theta_0 - \cos^2 \theta)}. \quad (108)$$

The constant $\alpha_1(0)/k^2$ is a pure number as the form of (108) shows. This number can be obtained by using the same sort of procedure as was used in the case of a half plane.

REFERENCES

- [1] B. B. Baker and E. T. Copson, "The Mathematical Theory of Huygens' Principle," Oxford University Press, London, Eng., 2nd ed.; 1950.
 - [2] F. C. Karal and S. N. Karp, "Diffraction of a skew plane electromagnetic wave by an absorbing right-angled wedge," *Commun. Pure Appl. Math.*, vol. 11, pp. 495-533; November, 1958.
 - [3] F. C. Karal and S. N. Karp, "Diffraction of a Plane Wave by a Right Angled Wedge Which Sustains Surface Waves on One Mace," Div. EM Res., Inst. Mathematical Science, New York University, New York, N. Y., Res. Rept. No. EM-123; January, 1959. To be submitted for publication.
 - [4] S. N. Karp, "Two Dimensional Green's Function for a Right-Angled-Wedge Under an Impedance Boundary Condition," Div. EM Res., Inst. of Mathematical Science, New York University, New York, N. Y., Res. Rept. No. 129; March, 1959. To be submitted for publication.
 - [5] S. N. Karp and F. C. Karal, "Surface Waves on a Right-Angled Wedge," Div. EM Res., Inst. of Mathematical Science, New York University, New York, N. Y., Res. Rept. No. EM-116; August, 1958. Condensed version in 1958 IRE WESCON CONVENTION RECORD, Pt. I, pp. 101-103. To appear shortly in *Commun. Pure Appl. Math.*
 - [6] S. N. Karp and F. C. Karal, "Vertex Excited Surface Waves on One Face of a Right Angled Wedge," N.Y.U., Inst. of Mathematical Science, Div. EM Res., New York University, New York, N. Y., Res. Rept. No. EM-124; January, 1959. To appear shortly in *Quart. Appl. Math.*
 - [7] H. Lamb, "On Sommerfeld's diffraction problem and on reflection by a parabolic mirror," *Proc. London Math. Soc.*, series 2, vol. 4, pp. 190-203; 1906.
 - [8] H. Lewy, "Waves on sloping beaches," *Bull. Amer. Math. Soc.*, vol. 52, pp. 737-775; 1946.
 - [9] J. J. Stoker, "Surface waves in water of variable depth," *Quart. Appl. Math.*, vol. 5, pp. 1-54; 1947.
-

Some New Forms of Huygens' Principle*

V. H. RUMSEY†

Summary—Several theorems are derived by using the reaction concept. They apply to single frequency sources of finite extent, and to fields which are finite and continuous on a hypothetical closed surface S whose inside or outside is source-free.

1) The field on the source-free side of S is uniquely determined by the normal components E_n and H_n on S .

2) For measurements on the source-free side of S , the primary source can be replaced by any of the following secondary sources on S :

- (a) Normal electric and magnetic dipoles in free space.
- (b) Normal electric and magnetic quadrupoles backed by a medium which makes $E_n=0$ and $H_n=0$.
- (c) A certain combination of electric and magnetic normal dipoles and quadrupoles which gives zero field on the source side of S .

The prescription for (a) requires the solution for the field of the primary source when it is inside a cavity on whose walls $E_n=0$ and $H_n=0$. The surface densities of the various components of secondary sources (b) and (c) are given in terms of the free-space primary field. The formulas are comparatively simple when S is plane.

Many applications are cited including a combination of multipoles, consisting of vertical electric and magnetic dipoles, which is equivalent to a horizontal electric dipole.

I. INTRODUCTION

HUYGENS proposed his famous principle in connection with optical effects. At the time, the electromagnetic nature of light was unknown, and consequently the original statement of the principle appears rather vague to the present-day reader. However, it turns out that the idea behind Huygens' principle is confirmed by field theory and thus applies not only to optics but essentially to any part of physics which can be treated by some type of field theory, such as gravitation, elasticity, acoustics, electricity, magnetism and many more. We therefore begin with a statement of the principle which applies to any of these phenomena, as follows.

Given some source denoted by g , we imagine a hypothetical surface S which encloses g . Then there is a certain source h which when spread over S gives the same field outside S as g . In other words, h is equivalent to g for all points of observation outside S . The equivalent or Huygens' source h is not unique: the different forms of h give different internal fields but all give the same external field of course.

Although this is intended to be a theoretical study, the widespread practical applications of Huygens' idea should be mentioned. It will be seen from our statement of the principle that the first step in any application is to pick some closed surface S which separates the

active part of the system (the source) from the passive part. As an illustration, consider the problem of designing an aircraft antenna for a certain radiation pattern, for example the pattern of an elementary dipole. Anyone who has made measurements of aircraft antennas would soon become convinced of the utter impossibility of doing this at, say, 10-cm wavelength, because the pattern of a single antenna on the aircraft is very complicated and adding more antennas seems to make it worse. Yet a solution exists, and Huygens' principle tells us what it is: taking S as the surface of the aircraft, which we assume to be closed and perfectly conducting, S must be covered with tangential magnetic dipoles of surface density $K = E \times n$ where E is the field of the dipole whose field is to be simulated. The current interest in stereophonic sound suggests another example. In this case the active part of the system is the orchestra; so we can take S as an imaginary cubicle surrounding one of the audience. The next step is to construct an actual model of this cubicle making sure the inside is the same as it was in the orchestra hall. Then the listener seated in the cubicle will hear the orchestra exactly as it was in the hall, if the walls of the cubicle are covered with loudspeakers hooked up so as to reproduce the original distribution of normal velocity or pressure over S . Again, to produce an even 70°F in a room, the walls must be covered with heaters (or coolers) so as to maintain 70° at the walls. The temperature anywhere inside will be exactly 70° no matter how hot or cold it is outside. The problem of reproducing a picture in three dimensions, which is of current interest to the film industry, is another example. If the screen surrounded the audience we would have our closed surface S which can now be thought of as a window separating the viewers from the picture. Then we would get a result which is optically indistinguishable from the real thing if we could project onto S the same distribution of light as in the real case. There is no end to the applications, and, in the light of practical experience, the results predicted by the principle are in many cases amazing.

The literature on Huygens' principle [1]–[8] goes back to 1690, but the material which refers to electromagnetic waves is practically confined to the twentieth century. It is a rather difficult subject and the reader will find considerable differences between the references cited.

The principle is probably most familiar to radio engineers under the name of Schelkunoff's equivalence theorem although Schelkunoff pointed out that Larmor originated this particular theorem. On the other hand, mathematicians would probably think of the subject as a collection of theorems which state how the field on one side of a closed surface is uniquely determined by its behavior on the boundary.

* This work was supported in part by the Office of Naval Research, under Contract No. N7onr-29529.

† Electronics Res. Lab., University of California, Berkeley, Calif.

In the present paper, the subject will be approached from the point of view of the reaction concept [9]. Since this method is probably unfamiliar to many readers, the reaction approach will first be explained by showing how it is used to derive some of the established forms of Huygens' principle. The method will then be used to obtain some new results.

Turning now to the specific case of radio waves, suppose that the medium is uniform, *e.g.*, free space, and that S is an infinite plane (which is not a closed surface but, as we shall see, the principle still applies). Then some of the established forms of Huygens' principle are as follows:

1) \mathbf{h} consists of tangential electric and magnetic dipole distributions of surface densities J_h and K_h where

$$J_h = \mathbf{n} \times \mathbf{H}_g \quad \text{and} \quad K_h = \mathbf{E}_g \times \mathbf{n}, \quad (1)$$

\mathbf{E}_g and \mathbf{H}_g denoting the electric and magnetic fields at S due to \mathbf{g} (*i.e.*, the primary field) and \mathbf{n} denoting a unit vector normal to S pointing away from \mathbf{g} ;

2) \mathbf{h} is a tangential magnetic dipole distribution of surface density $2K_h$;

3) \mathbf{h} is a tangential electric dipole distribution of surface density $2J_h$.

The first form gives zero field on the left, whereas 2) and 3) plainly do not. Note also that 1) requires a knowledge of both tangential \mathbf{E}_g and tangential \mathbf{H}_g , whereas 2) or 3) require either tangential \mathbf{E}_g or tangential \mathbf{H}_g , respectively.

Alternatively, using the Hertz potential method [10]–[17], we would express the field to the right of S in the form

$$\mathbf{E}_g = \mathbf{E}_e + \mathbf{E}_m \quad \text{and} \quad \mathbf{H}_g = \mathbf{H}_e + \mathbf{H}_m \quad (2)$$

where

$$\mathbf{E}_m = \nabla \times \mathbf{n} f_m$$

and from Maxwell's equation

$$\begin{aligned} -Z\mathbf{H}_m &= \nabla \times \mathbf{E}_m, \\ \mathbf{H}_e &= \nabla \times \mathbf{n} f_e \end{aligned}$$

and from Maxwell's equation

$$Y\mathbf{E}_e = \nabla \times \mathbf{H}_e, \quad (3)$$

the scalar functions f_m and f_e being the Hertz potentials. In words, (2) and (3) state that the field is represented as the superposition of a transverse electric (TE) part ($\mathbf{E}_e\mathbf{H}_e$) and a transverse magnetic (TM) part ($\mathbf{E}_m\mathbf{H}_m$), TE and TM being defined with respect to the \mathbf{n} axis. Note that (3) suggests that \mathbf{E}_e and \mathbf{E}_m are the fields of certain distributions of electric and magnetic dipoles pointing in the direction normal to S .

Now let us compare these two ways of representing the field to the right of S . First note that form 3) of Huygens' principle is similar to the Hertz potential method (3) in this respect; they both express the field in terms of two complex scalar functions, the former in

terms of the tangential components of \mathbf{H}_g at S and the latter in terms of f_m and f_e . They differ in the respect that the equivalent source in 3) consists of tangential electric dipoles, but in (3) it consists of normal electric and magnetic dipoles. This suggests that if (3) is valid in general we ought to be able to find a form of Huygens' principle in which \mathbf{h} consists of normal electric and magnetic dipoles. The study of such a theorem for an arbitrary surface is the main purpose of this paper.

II. THE REACTION METHOD

Maxwell's equations for a single frequency can be written in the phasor form

$$-\nabla \times \mathbf{E} = Z\mathbf{H} + \mathbf{K} \quad Z = j\omega\mu + \tau \quad (4)$$

$$\nabla \times \mathbf{H} = Y\mathbf{E} + \mathbf{J} \quad Y = j\omega\epsilon + \sigma \quad (5)$$

in which \mathbf{E} and \mathbf{H} represent the radio wave field, Z and Y the environment, \mathbf{J} and \mathbf{K} the source; \mathbf{E} , \mathbf{H} , \mathbf{J} , and \mathbf{K} are complex vectors. The source functions \mathbf{J} and \mathbf{K} represent the densities of electric and magnetic current moment, or let us say simply the electric and magnetic dipoles. Let $J_a K_a$ and $J_b K_b$ denote two different sources, $\mathbf{E}_a \mathbf{H}_a$ and $\mathbf{E}_b \mathbf{H}_b$ denoting the corresponding fields. Let

$$\langle ab \rangle = \iiint_{\text{over } a} (\mathbf{J}_a \cdot \mathbf{E}_b - \mathbf{K}_a \cdot \mathbf{H}_b) dV, \quad (6)$$

the integral being taken over all space, *i.e.*, over the source \mathbf{a} , since there is no contribution where \mathbf{J}_a and \mathbf{K}_a vanish. Then it can be shown that

$$\langle ab \rangle = \langle ba \rangle \quad (7)$$

$$\langle ab \rangle = \oint_S (\mathbf{E}_a \times \mathbf{H}_b - \mathbf{E}_b \times \mathbf{H}_a) \cdot \mathbf{n} dS \quad (8)$$

with these provisions.

- 1) The sources do not extend to infinity or, more precisely, the integral in (6) is finite.
- 2) Z and Y are scalars (isotropic medium).
- 3) There is some loss (or radiation).
- 4) In (7), $Z_a = Z_b$, $Y_a = Y_b$ everywhere; *i.e.*, the sources \mathbf{a} and \mathbf{b} are in the same environment.
- 5) S separates \mathbf{a} and \mathbf{b} ; *i.e.*, on one side of S $J_a = 0$, $K_a = 0$ and on the other side $J_b = 0$, $K_b = 0$. The unit normal vector \mathbf{n} points from \mathbf{a} to \mathbf{b} .
- 6) In (8), the environment for \mathbf{b} is the same as in (6), but for \mathbf{a} it is the same only on the side of S where \mathbf{a} is situated; it can be different on the other side (obviously \mathbf{a} and \mathbf{b} can be interchanged in this statement).

If \mathbf{a} and \mathbf{b} are on the same side of S

$$0 = \oint_S (\mathbf{E}_a \times \mathbf{H}_b - \mathbf{E}_b \times \mathbf{H}_a) \cdot \mathbf{n} dS \quad (9)$$

provided that $Z_a = Z_b$ and $Y_a = Y_b$ on the other side of S . The complex number $\langle ab \rangle$ is called the reaction of \mathbf{a}

on \mathbf{b} . It continually arises in theoretical work and so it is convenient to give it a name, "reaction," and a simple notation, $\langle ab \rangle$. It can be thought of as a fundamental observable which represents the coupling between sources \mathbf{a} and \mathbf{b} ; e.g., if \mathbf{b} represents a unit current generator connected to an antenna, $\langle ab \rangle$ is equal to the voltage it receives from \mathbf{a} . The property which we shall use frequently is that if

$$\langle at \rangle = \langle bt \rangle$$

for all test sources \mathbf{t} , then sources \mathbf{a} and \mathbf{b} are equivalent; e.g., $E_a = E_b$ and $H_a = H_b$ everywhere. This follows immediately by taking \mathbf{t} as an electric dipole of unit moment so that, by (6), $\langle at \rangle$ becomes the component of E_a parallel to the dipole moment at the location of the dipole.

As an example of the reaction method, take \mathbf{a} in (8) as the primary or given source \mathbf{g} in the discussion of Huygens' principle and S as the hypothetical surface which encloses \mathbf{g} . Take \mathbf{b} as a test source \mathbf{t} located somewhere outside S . Then, with a minor rearrangement of the integrand, (8) takes the form

$$\langle gt \rangle = \oint_S [(n \times H_g) \cdot E_t - (E_g \times n) \cdot H_t] dS. \quad (10)$$

But, from (6) we see that if \mathbf{h} consists of the surface densities J_h and K_h of electric and magnetic dipoles on S , by definition

$$\langle ht \rangle = \oint_S [J_h \cdot E_t - K_h \cdot H_t] dS. \quad (11)$$

Therefore if

$$J_h = n \times H_g \quad \text{and} \quad K_h = E_g \times n, \\ \langle gt \rangle = \langle ht \rangle$$

for all \mathbf{t} outside S .

In other words, the Huygens' source \mathbf{h} is equivalent to the given source. Similarly by taking \mathbf{t} inside S , (8) shows that $\langle ht \rangle = 0$; i.e., \mathbf{h} has zero effect inside S . Thus we have established form (1) of Huygens' principle, mentioned earlier. Note that the derivation is valid for the general heterogeneous isotropic medium: it can be extended to the anisotropic case by using a more general formula for the reaction [9].

A quicker way of getting this result is to argue as follows. Consider the field which is equal to $E_g H_g$ outside S and zero inside S . Then the boundary conditions at S require the electric and magnetic dipole distributions given by (11). However, this kind of argument assumes that such a field can exist, whereas the method preceding (11) does not depend on this assumption.

To show that S can be replaced by an infinite plane, take S as the surface of a hemisphere consisting of S_P the plane part and S_C the curved part. Then we have to show that the contribution from \mathbf{k} , the part of \mathbf{h} on S_C , becomes insignificant for large radius. Thus, consider

$$\begin{aligned} \langle kt \rangle &= \iint_{S_C} (J_k \cdot E_t - K_k \cdot H_t) dS && \text{by definition} \\ &= \iint_{S_C} (n \times H_g \cdot E_t - E_g \times n \cdot H_t) dS && \text{from (11).} \end{aligned}$$

But at large distances R from \mathbf{g} and \mathbf{t}

$$\begin{aligned} n \times E_g &= \sqrt{\frac{Z}{Y}} H_g + 0 \left[\frac{1}{R^2} \right] && (n \text{ being radial}) \\ n \times E_t &= \sqrt{\frac{Z}{Y}} H_t + 0 \left[\frac{1}{R^2} \right] \end{aligned}$$

assuming a uniform medium at large R . Therefore,

$$\langle kt \rangle \sim 0 \left[\frac{1}{R} \right] \rightarrow 0$$

as was to be proved.

The other forms of Huygens' principle mentioned in the introduction are obtained by using condition 6). For ease of reference, let us take \mathbf{b} inside S (and \mathbf{a} outside therefore). Put $Y_b = \infty$ outside S : this is allowed by condition (6). Thus, \mathbf{b} is in cavity with $n \times E_b = 0$ on the wall S , but \mathbf{a} is in free space (for example). The field of \mathbf{b} is therefore not the same as in free space and to distinguish it we use the subscript 1. Thus, $E_{b1} H_{b1}$ represents the field of \mathbf{b} in the cavity. Then (8) gives

$$\langle ba \rangle = \oint_S n \times H_{b1} \cdot E_a dS = \oint_S E_a \times n \cdot H_{b1} dS \quad (12)$$

where \mathbf{n} points away from \mathbf{b} .

Consider \mathbf{b} as the test source in this relation. Then if \mathbf{c} represents the surface density of magnetic dipoles K_c where

$$K_c = E_a \times n, \text{ where } \mathbf{n} \text{ points away from } \mathbf{a}; \quad (13)$$

$$\langle ba \rangle = \langle bc \rangle_1 \text{ for any } \mathbf{b} \text{ inside } S. \quad (14)$$

Note that $\langle ba \rangle$ refers to free space and $\langle bc \rangle_1$ refers to the cavity. The physical interpretation of (14) is, therefore, that \mathbf{c} in the cavity gives the same field as \mathbf{a} in free space. Thus, we obtain a Huygens source consisting of magnetic dipoles $K = E \times n$ backed by a perfect conductor ($Y = \infty$). (This result can be obtained more directly as follows. It is known that tangential E is discontinuous at a tangential magnetic dipole sheet K by the amount $n \times K$. Therefore if one side of the sheet is a perfect conductor, on the other side tangential E must equal $n \times K$. Thus, the magnetic dipole sheet $K = E_g \times n$, backed by a perfect conductor, reproduces the same tangential E on S as E_g . However, this argument assumes that tangential E on S specifies the field uniquely, which can be established by a similar argument [18], [19].) When S is an infinite plane and the medium is uniform, the method of images shows that K backed by $Y = \infty$ is equivalent to $2K$ in free space [18], [19], which gives form 2) of Huygens' principle mentioned in the Introduction.

Returning to (12) now consider \mathbf{a} as the test source. Then if \mathbf{d} represents the surface density of electric dipoles

$$J_d = n \times H_{b1}, \text{ where } n \text{ points away from } \mathbf{a}; \quad (15)$$

$$\langle ba \rangle = \langle da \rangle \text{ for all } a. \quad (16)$$

Hence, \mathbf{d} is equivalent to \mathbf{b} : the Huygens source consists of the electric dipoles $J_d = n \times H_{b1}$ in free space. Now since $E_{b1}H_{b1}$ is the field of \mathbf{a} in a conducting cavity, minus J_d is the current induced in the wall. (This form of Huygens' principle can also be deduced from more elementary considerations for, assuming that the total field beyond the wall is zero, the field of the wall currents must be equal and opposite to the free-space field of \mathbf{a} for all external points.) The form 3) of Huygens' principle mentioned in the Introduction is obtained by taking S as an infinite plane and using the method of images. This gives $n \times H_{a1} = 2n \times H_a$ and, thus, we obtain form 3). Note that "inside" and "outside" can be interchanged; *i.e.*, these forms of Huygens' principle apply to internal or external regions.

It may seem that the reaction method is a rather heavy handed way of getting these results because they can be derived by using simpler methods. There are two objections to these simpler methods, however. One is that they are not rigorous; the other is that they cannot be extended to the case of normal dipoles.

Analogous results are obtained by a like modification of Z instead of Y (or by taking $Y=0$ instead of $Y=\infty$). More generally, by taking $Z=\infty$ over part of S and $Y=\infty$ over the rest, we obtain \mathbf{h} in the form of tangential electric dipoles over part of S and tangential magnetic dipoles over the rest. This leads to the theorem that the field is uniquely determined by the boundary values of tangential E over part of S and tangential H over the rest, as follows. If \mathbf{g}_1 and \mathbf{g}_2 are two sources having these boundary conditions in common, then it follows immediately that $\langle g_1 t \rangle = \langle g_2 t \rangle$ for all t ; *i.e.*, \mathbf{g}_1 and \mathbf{g}_2 are indistinguishable on the source-free side of the boundary.

Although the condition $n \times E = 0$ at a perfect conductor is well known, it is seldom well proved in the text books, and since this is clearly a key point, some comment may be worth while. To prove the point it must be deduced from the Maxwell equations (4) and (5) by considering the limit as $Y \rightarrow \infty$ on one side of a surface S . One way of doing this is to show that the intrinsic wavelength $\lambda \rightarrow 0$ as $Y \rightarrow \infty$. Then, relative to λ , S is effectively plane, provided its radius of curvature is finite. Hence, the problem is reduced to consideration of a plane boundary which can be analyzed fairly simply.

In summary, the method described in this section consists of the following steps:

1) Apply some appropriate boundary condition on the field of source \mathbf{b} to the reaction integral (8).

2) Manipulate the result so that it has the standard form for the reaction with source \mathbf{a} . Treat \mathbf{a} as a test source to get an equivalence theorem for the source \mathbf{b} .

3) Treat \mathbf{b} as a test source in this relation to get a uniqueness theorem in terms of the boundary values of E_a and H_a .

4) Manipulate the result of 1) so that it has the standard form for the reaction with source \mathbf{b} . Treat \mathbf{b} as a test source to get a different type of equivalence theorem.

In the following sections we will carry out these steps for the boundary condition $n \cdot E_b = 0$ and $n \cdot H_b = 0$.

III. THE BOUNDARY CONDITION $n \cdot E = 0$, $n \cdot H = 0$

The preceding section suggests that to investigate whether the field can be uniquely represented in terms of $n \cdot E$ and $n \cdot H$ at S , we should consider a sheet of sources which gives a prescribed discontinuity in $n \cdot E$ and $n \cdot H$ backed with a medium which makes $n \cdot E = 0$ and $n \cdot H = 0$. In this section we consider the possibility of such a medium, the main point being to show that the imposition of the boundary conditions $n \cdot E = 0$ and $n \cdot H = 0$ on S is consistent with the Maxwell equations (4) and (5), and is similar to the more familiar boundary conditions at a perfect conductor.

It follows from (4) and (5) that if $J=0$ and $K=0$, the normal components of YE and ZH are continuous, *i.e.*,

$$n \cdot Y_0 E_0 = n \cdot YE$$

and

$$n \cdot Z_0 H_0 = n \cdot ZH,$$

the subscript zero denoting the free-space side of the boundary. Taking Y_0 and Z_0 to be finite scalars, it follows that if $n \cdot E_0 = 0$ and $n \cdot H_0 = 0$, then $n \cdot YE = 0$ and $n \cdot ZH = 0$. If we take Y and Z to be tensors, these equations do not necessarily imply $Y=0$ and $Z=0$; we have $E \cdot \tilde{Y}n = 0$ and $H \cdot \tilde{Z}n = 0$, where \tilde{Y} denotes the transpose of Y . We see that these equations can be satisfied for any E and H provided $\tilde{Y}n = 0$ and $\tilde{Z}n = 0$, giving three restrictions on the nine components of Y or Z . For example, if n represents the z coordinate, $Y_{zz} = 0$, $Y_{zy} = 0$ and $Y_{zx} = 0$, but the first two rows of Y are arbitrary. The medium which makes $n \cdot E = 0$ and $n \cdot H = 0$ at its boundary is therefore not unique, but then neither is a "perfect conductor" unique because Z is arbitrary.

By considering the problem of an arbitrary plane-wave incident on the plane boundary $z=0$ of material with $Z_{ij}=0$ and $Y_{ij}=0$ except the diagonal terms Z_{11} , Z_{22} , Z_{33} , Y_{11} , Y_{22} , Y_{33} , it is easily verified that the propagation constant in the normal (or z) direction tends to infinity as $Z_{33} \rightarrow 0$ and $Y_{33} \rightarrow 0$ regardless of Z_{11} , Z_{22} , Y_{11} , and Y_{22} , except for normal incidence. (The solution of the problem for the case of waves traveling down a waveguide is given in the Appendix. It is found that the propagation constant tends to infinity for every mode, and therefore the transmitted field dies out with infinite rapidity for any incident field.) The solution for an arbitrary source above such a medium is given explicitly in Section VIII. The propagation within the medium thus meets the conditions for geometrical optics and a

curved boundary can therefore be treated as locally plane. That normal incidence is exceptional is to be expected because $E \cdot n$ and $H \cdot n$ then vanish everywhere, and obviously the imposition of the boundary condition $E \cdot n = 0$ and $H \cdot n = 0$ has no effect. With this exception (which applies only to the infinite plane boundary) we see that the medium is effectively unique in the same sense that a perfect conductor is effectively unique. (Section VIII shows that this exception is insignificant.)

We shall now show that if $n \cdot E = 0$ and $n \cdot H = 0$ on a closed surface S , then

$$n \times E = n \times \nabla \phi \quad \text{and} \quad n \times H = n \times \nabla \psi; \quad (17)$$

i.e., the tangential components of E or H can be expressed as the gradient of a scalar. To prove (17) we take a curvilinear coordinate system in which n forms the first unit vector. If the external medium is isotropic, it follows from $n \cdot H = 0$ that

$$\begin{aligned} n \cdot \nabla \times E &= (\nabla \times E)_1 \\ &= \frac{1}{h_2 h_3} \left[\frac{\partial}{\partial u_2} (h_3 E_3) - \frac{\partial}{\partial u_3} (h_2 E_2) \right] = 0 \\ \therefore h_3 E_3 &= \frac{\partial \phi}{\partial u_3} \quad \text{and} \quad h_2 E_2 = \frac{\partial \phi}{\partial u_2} \end{aligned}$$

which is equivalent to the first of (17). The second follows in like manner.

IV. FORMULAS IN TERMS OF $n \cdot E$ AND $n \cdot H$

We now apply the boundary conditions $n \cdot E_b = 0$ and $n \cdot H_b = 0$ to the formula

$$\langle ab \rangle = \oint_S (E_a \times H_b - E_b \times H_a) \cdot n dS. \quad (18)$$

Then, since $\langle ab \rangle$ involves only the tangential components of E_b and H_b , according to (17) we can replace E_b by $\nabla \phi_b$ and H_b by $\nabla \psi_b$. Now the term

$$\begin{aligned} E_a \times H_b \cdot n &= H_a \times \nabla \psi_b \cdot n \\ &= (\psi_b \nabla \times E_a - \nabla \times \psi_b E_a) \cdot n \\ &= (-\psi_b \nabla H_a - \nabla \times \psi_b E_a) \cdot n. \end{aligned}$$

Therefore,

$$\oint_S E_a \times H_b \cdot n dS = - \oint_S \psi_b \nabla H_a \cdot n dS, \quad (19)$$

assuming ψ_b is single valued. Similarly,

$$\oint_S H_a \times E_b \cdot n dS = - \oint_S Y \phi_b \nabla E_a \cdot n dS. \quad (20)$$

Substitution of (19) and (20) in (18) gives

$$\langle ab \rangle = \oint_S (Y \phi_b n \cdot E_a - Z \psi_b n \cdot H_a) dS, \quad (21)$$

which expresses the field of a in terms of $n \cdot E_a$ and $n \cdot H_a$ on S . It shows that the field on the source-free side of S is uniquely determined by $n \cdot E$ and $n \cdot H$ at S , for if a_1

and a_2 represent two sources which have the same boundary values of $n \cdot E$ and $n \cdot H$, (21) shows that $\langle (a_1 - a_2) b \rangle = 0$ for all b on the source-free side of the boundary. It will be noted that ϕ_b and ψ_b are not unique because they are defined by $n \times E_b = n \times \nabla \phi_b$ and $n \times H_b = n \times \nabla \psi_b$. That this makes no difference to the surface integral in (21) is easily verified. Corresponding to this flexibility in the choice of ϕ_b and ψ_b there are the restrictions

$$\oint_S Y n \cdot E_a dS = 0 \quad \text{and} \quad \oint_S Z n \cdot H_a dS = 0 \quad (22)$$

on $n \cdot E_a$ and $n \cdot H_a$ which follow from the Maxwell equations (4) and (5).

As a simple application we can use (21) to prove the impossibility of an isotropic source. If there were such a thing, it would give uniform $n \cdot E$ and $n \cdot H$ on the surface of a sphere. But, as we have just seen, this is inconsistent with Maxwell's equations. Therefore, $n \cdot E$ and $n \cdot H$ must be zero and (21) shows that this means the field must be zero.

The physical interpretation of (21) is seen by comparing it with the fundamental formula for reaction (6). Eq. (21) states that b is equivalent to a surface density $\phi_b Y$ of normal electric dipoles and a surface density $\psi_b Z$ of normal magnetic dipoles in free space. Note that ϕ_b and ψ_b are determined by the tangential components of E_b and H_b at the boundary which makes $n \cdot E_b = 0$ and $n \cdot H_b = 0$. Thus, to find the equivalent density of normal dipoles, we first imagine source b to be inside a cavity in such a material as to make $n \cdot E_b = 0$ and $n \cdot H_b = 0$ at the cavity wall. Then assuming the field beyond the cavity wall to be zero, which can be justified by considering the plane boundary in the manner indicated in Section III (see the Appendix), the field due to the sources induced in the wall must be equal and opposite to the field of b . The induced sources must therefore consist of the normal electric and magnetic dipoles $-\phi_b Y$ and $-\psi_b Z$ (a result which is verified directly in the Appendix). The determination of this form of Huygens' source requires the solution to the cavity problem described above and evidently can be carried out only when S is a particularly simple surface. For example, if S is the cross section of a lossless waveguide (S is then effectively closed) and the incident field is a TE mode, the normal dipole distribution induced on the surface of the anisotropic boundary turns out to be proportional to the longitudinal component of H in the incident mode (see the Appendix). In this case (21) states that the field radiated from the open end of such a waveguide is the same as if we put this distribution of longitudinal magnetic dipoles in the opening and removed the primary source.

There are some problems for which ϕ and ψ are not single valued. For example take a coaxial line made of perfect conductors. Suppose that the field is the TEM mode (for which $E_n = 0$ and $H_n = 0$, n being in the direction of the axis). Then if we choose S to be normal to the z axis in between the conductors, as illustrated in Fig. 1,

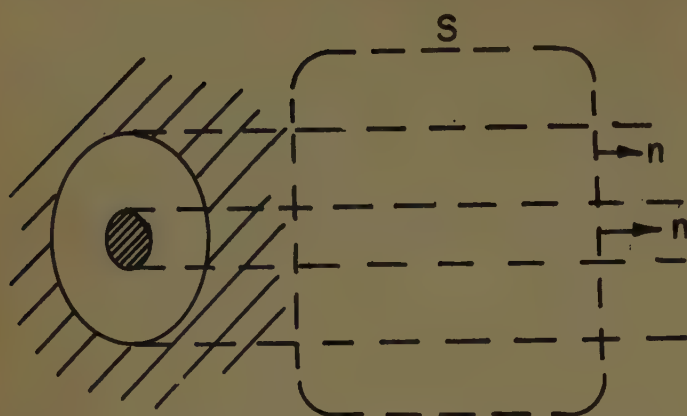


Fig. 1—The closed surface S is chosen so that, at every point between the coaxial conductors, the normal n is parallel to the z axis.

the normal components of E and H vanish everywhere on S , or so it seems. The vanishing on the part of S between the conductors is due to the choice of S , and the vanishing on the rest is due to the fact that the field in a perfect conductor vanishes. If this be so, it is clear that the theorem does not hold, because here we have a nonvanishing field with $n \cdot E = 0$ and $n \cdot H = 0$ on S . Let us turn to the function ψ introduced in (17). We imagine our artificial medium, which makes $E_z = 0$ and $H_z = 0$ at its surface, to be placed just outside of S . But since $E_z = 0$ and $H_z = 0$ everywhere, this has no effect. Thus the E and H in (17) are apparently the actual values. Then

$$\oint_C \nabla \psi \cdot d\mathbf{l} = \oint_C \mathbf{H} \cdot d\mathbf{l} = I$$

the current on the inner conductor, if C encircles the inner conductor. This shows that ψ is not single-valued and consequently the proof of the theorem does not apply.

While this argument is consistent, it is open to question because the field has a singularity just inside the inner conductor; the vector $\nabla \times \mathbf{H}$ behaves like a Dirac δ -function and therefore the equation $n \cdot \nabla \times \mathbf{H} = 0$ is meaningless. In effect, we did not treat S as a closed surface, but as two parts, without considering what happens on the boundary line between these two parts. The problem of the singularity can be approached by starting with conductors which are not perfect and taking the limit as $\sigma \rightarrow \infty$. Now the boundary condition $n \cdot H = 0$ $n \cdot E = 0$ is no longer without effect because the TEM mode cannot exist for finite conductivity. In this case, it is easy to check that the theorem applies without question. The result for the perfect conductor thus depends on whether we introduce the boundary condition $n \cdot E = 0$ $n \cdot H = 0$ before, or after, taking the limit $\sigma \rightarrow \infty$.

In summary, (21) states that any source can be replaced by a distribution of normal electric and magnetic poles spread over the closed surface S , provided the field is finite and continuous on S . (The equivalence applies only on the source-free side of S , of course). It

also shows that the field is uniquely determined by the boundary values of $n \cdot E$ and $n \cdot H$ subject to the same provision.

V. NORMALLY DIRECTED QUADRUPOLES

We have seen that by considering a as a test source, (21) gives a form of Huygens' principle in which the Huygens source consists of normally directed dipoles in free space. Alternatively, we can consider b as a test source but then the right side of (21) is not in a standard form so that the physical interpretation is not clear. It turns out that (21) can be put into the form of the reaction with a certain quadrupole distribution. We therefore take up the properties of quadrupoles in this section.

The general quadrupole can be taken as two dipoles of moment $\pm p$ separated by an amount represented by the vector m , as illustrated in Fig. 2. Thus, if q represents an electric quadrupole and t a test source,

$$\langle qt \rangle = E_t(A) \cdot p - E_t(B) \cdot p,$$

where $E_t(A)$ = field of t at point A .

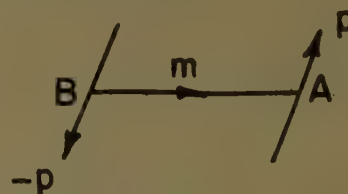


Fig. 2—A quadrupole consisting of two dipoles.

Thus, as A tends to coincide with B ,

$$\langle qt \rangle = \delta x \frac{\partial(E_t \cdot p)}{\partial x} + \delta y \frac{\partial(E_t \cdot p)}{\partial y} + \delta z \frac{\partial(E_t \cdot p)}{\partial z},$$

where $\delta x, \delta y, \delta z$ are the components of m . Therefore,

$$\langle qt \rangle = m \cdot \nabla(E_t \cdot p).$$

If p and m are both parallel to n ,

$$\langle qt \rangle = Q n \cdot \nabla E_{tn},$$

where Q represents the strength of this normally directed quadrupole. Thus, for a surface density f_q of such quadrupoles

$$\langle qt \rangle = \oint_{\text{over } f_q} f_q n \cdot \nabla E_{tn} dS. \quad (23)$$

Take a system of curvilinear coordinates with n as the first unit vector. Then in the conventional notation

$$n \cdot \nabla E_n = \frac{1}{h_1} \frac{\partial E_1}{\partial u_1} \quad (24)$$

(dropping the subscript t for the present).

Now $\nabla \cdot \mathbf{Y}E = 0$.

Therefore,

$$0 = \frac{1}{h_1} \frac{\partial Y E_1}{\partial u_1} + \frac{Y E_1}{h_1 h_2 h_3} \frac{\partial}{\partial u_1} (h_2 h_3) + \frac{1}{h_1 h_2 h_3} \left[\frac{\partial}{\partial u_2} (h_3 h_1 Y E_2) + \frac{\partial}{\partial u_3} (h_1 h_2 Y E_3) \right]. \quad (25)$$

Note that

$$n \cdot \nabla \times F = \frac{1}{h_2 h_3} \left[\frac{\partial}{\partial u_2} (h_3 F_3) - \frac{\partial}{\partial u_3} (h_2 F_2) \right].$$

Therefore, if

$$F = n \times h_1 Y E, \quad (26)$$

$$n \cdot \nabla \times F = \frac{1}{h_2 h_3} \left[\frac{\partial}{\partial u_2} (h_3 h_1 Y E_2) + \frac{\partial}{\partial u_3} (h_2 h_1 Y E_3) \right]. \quad (27)$$

Substitution in (25) gives

$$\begin{aligned} \frac{Y}{h_1} \frac{\partial E_1}{\partial u_1} + \frac{E_1}{h_1} \frac{\partial Y}{\partial u_1} \\ = \frac{1}{h_1} \frac{\partial Y E_1}{\partial u_1} = \frac{-Y E_1}{h_1 h_2 h_3} \frac{\partial}{\partial u_1} (h_2 h_3) - \frac{1}{h_1} n \cdot \nabla \times F \\ = -Y E_n \nabla \cdot n - \frac{n}{h_1} \cdot \nabla \times F. \end{aligned}$$

Therefore,

$$Y n \cdot \nabla E_n = -E_n \nabla \cdot n Y - \frac{n}{h_1} \cdot \nabla \times (n \times h_1 Y E). \quad (28)$$

We now introduce the boundary condition

$$E \cdot n = E_1 = 0. \quad (29)$$

Thus, the quadrupole layer q is on the surface of a material which makes normal E vanish. Then from (28) (29) and (23),

$$\langle q^t \rangle = - \iint_{\text{over } q} \frac{f_q}{Y} \frac{n}{h_1} \cdot \nabla \times F dS. \quad (30)$$

Now

$$\frac{f}{hY} \nabla \times F = \nabla \times \frac{fF}{hY} - \nabla \left(\frac{f}{hY} \right) \times F. \quad (31)$$

(This implies that f is a function of position in three dimensions, whereas f_a , when we introduced it in (23), was a function of position on S . To apply (31) in (30), we therefore take f at points on S equal to f_q .) Integration of the normal component over a closed surface eliminates the first term. Thus, (30) and (31), give,

$$\langle q^t \rangle = \iint_S \nabla \left(\frac{f}{hY} \right) \times F \cdot n. \quad (32)$$

now taking the surface to be closed. We now introduce the boundary condition

$$n \cdot H = 0,$$

which we saw gave

$$n \times E = n \times \nabla \phi$$

[see (17)]. Thus, f_a is now backed by the medium which makes $n \cdot E = 0$ and $n \cdot H = 0$. From (26), then,

$$F = Y n \times h \nabla \phi \quad (33)$$

$$\begin{aligned} \nabla \left(\frac{f}{hY} \right) \times F \cdot n &= \nabla \left(\frac{f}{hY} \right) \times (n \times Y h \nabla \phi) \cdot n \\ &= \left[n \times \nabla \left(\frac{f}{hY} \right) \right] \cdot n \times h \nabla \phi \\ &= n \cdot \nabla \phi \times \left[h Y n \times \nabla \left(\frac{f}{hY} \right) \right] \\ &= n \cdot \nabla \times \phi Y h \left[n \times \nabla \left(\frac{f}{hY} \right) \right] - \phi n \cdot \nabla \\ &\quad \times \left[h Y n \times \nabla \left(\frac{f}{hY} \right) \right]. \end{aligned}$$

Substitution in (32) eliminates the first term, giving

$$\langle q^t \rangle = - \oint_S \phi \cdot n \cdot \nabla \times \left[h_1 Y n \times \nabla \left(\frac{f_q}{h_1 Y} \right) \right]. \quad (34)$$

This has the same form as the first term in (21). Evidently the second term can be expressed in terms of the analogous magnetic quadrupole distribution. Hence, (21) gives a Huygens source consisting of surface densities f and g of normally-directed electric and magnetic quadrupoles backed by the medium which makes $n \cdot E = 0$ and $n \cdot H = 0$, f and g being given by

$$n \cdot \nabla \times \left[Y h_1 n \times \nabla \left(\frac{f}{h_1 Y} \right) \right] = -Y E_g \cdot n, \quad (35)$$

$$n \cdot \nabla \times \left[Z h_1 n \times \nabla \left(\frac{g}{h_1 Z} \right) \right] = -Z H_g \cdot n; \quad (36)$$

$n \cdot E_g$ and $n \cdot H_g$ being the normal values of the free-space given field. It will be noticed that (35) and (36) are unaffected by the addition of a constant to $f/h_1 Y$ for $g/h_1 Z$; so, like ϕ and ψ in Section IV, f and g are not unique. But again this does not lead to any lack of uniqueness in E or H because it follows from (28) that a quadrupole density $f = h_1 Y$ on a closed surface which makes $E_n = 0$, gives zero field.

Eqs. (35) and (36) can be written more concisely by introducing the surface operator ∇_T where

$$\nabla_T \cdot A = \frac{1}{h_1 h_2 h_3} \left[\frac{\partial (h_3 h_1 A_2)}{\partial u_2} + \frac{\partial (h_1 h_2 A_3)}{\partial u_3} \right],$$

$$\nabla_T V = \frac{u_2}{h_2} \frac{\partial V}{\partial u_2} + \frac{u_3}{h_3} \frac{\partial V}{\partial u_3}.$$

Then

$$\nabla_T \cdot Y \nabla_T \left(\frac{f}{h_1 Y} \right) = - \frac{Y E_{gn}}{h_1}, \quad (37)$$

$$\nabla_T \cdot Z \nabla_T \left(\frac{g}{h_1 Z} \right) = - \frac{Z H_{gn}}{h_1}. \quad (38)$$

For example, if the given field is the m th TM mode in a waveguide parallel to the z axis, taking S as a cross section of the waveguide gives

$$E_{gn} = E_{mz} \text{ and } H_{gn} = 0,$$

where

$$\nabla_T^2 E_{mz} = -K_m^2 E_{mz}$$

$$K_m = \text{mode number} = \frac{2\pi}{\text{cutoff wavelength}}.$$

In this case, $h_1 = 1$; and thus from (37) and (38),

$$f_m = \frac{Y E_{mz}}{K_m^2} \text{ and } g_m = 0. \quad (39)$$

Thus, the mode is generated by the quadrupole distribution $Y E_{mz}/K_m^2$ backed by the boundary which makes $n \cdot E = 0$ and $n \cdot H = 0$.

When the surface S is plane, the effect of the boundary conditions is easy to work out. Fig. 3 shows an electric dipole in front of a plane boundary on which $n \cdot E = 0$ and $n \cdot H = 0$. Since the dipole is in the normal direction, $n \cdot H = 0$ everywhere. We see that an equal and opposite image gives $n \cdot E = 0$. Fig. 4 shows what this result gives for a quadrupole: we see that the image of the quadrupole is the same as the object. When the quadrupole is on the surface it is therefore the same as one of double strength in free space. Thus, in our example of the previous paragraph, the mode is generated by the quadrupole density $2 Y E_{mz}/K_m^2$ spread over the cross section of the waveguide.

By the same kind of argument, it is easily seen that a normal magnetic quadrupole on an ideal metal plane ($n \times E = 0$) is equivalent to one of double strength in free space. Thus, for such a source there is no difference between the boundary conditions $n \cdot E = 0$ and $n \cdot H = 0$ and the boundary condition $n \times E = 0$. Now consider a circular waveguide connected to an aperture in a metal plane so that its open end forms the aperture. If only TE modes are present, the argument of the previous paragraph shows that the radiating aperture is equivalent to a distribution of magnetic quadrupoles in free space. Thus, we can construct a practical magnetic quadrupole antenna by exciting the waveguide with the lowest TE mode for which E is circular and uniform around the axis. From symmetry it can be seen that the diffraction at the aperture will not generate any TM modes, and, thus, the condition for pure magnetic quadrupole radiation is satisfied. The electric field everywhere is parallel to the metal plane. To simulate a point quadrupole, the diameter of the waveguide would have to be small compared to the wavelength, which would require dielectric loading of the waveguide.

Fig. 5 shows how dipole and quadrupole antennas can be constructed by using shunt or series excitation. In practice they will give a mixture of the various electric and magnetic dipole and quadrupole effects; to avoid this, l/d and λ/l should be as large as possible. The electric antenna is coaxial. The magnetic antenna is a split cylinder, which in the case of the quadrupole is bisected except for a thin connection on one side of the split.

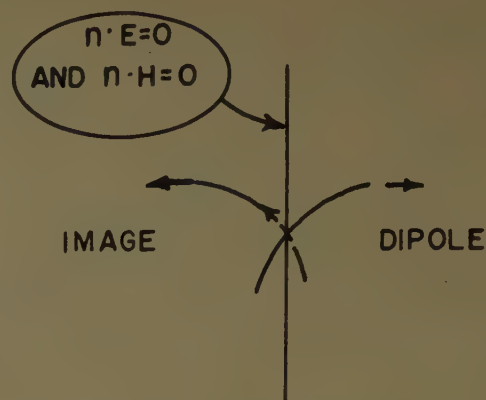


Fig. 3—The image of a normal dipole in a surface where $n \cdot E = 0$ and $n \cdot H = 0$ is out of phase with the dipole.

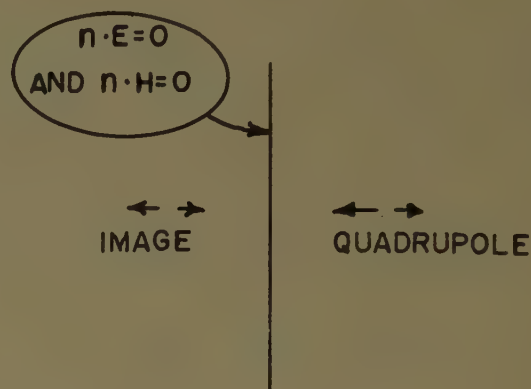


Fig. 4—The image of a normal quadrupole in a surface where $n \cdot E = 0$ and $n \cdot H = 0$ is in phase with the quadrupole.

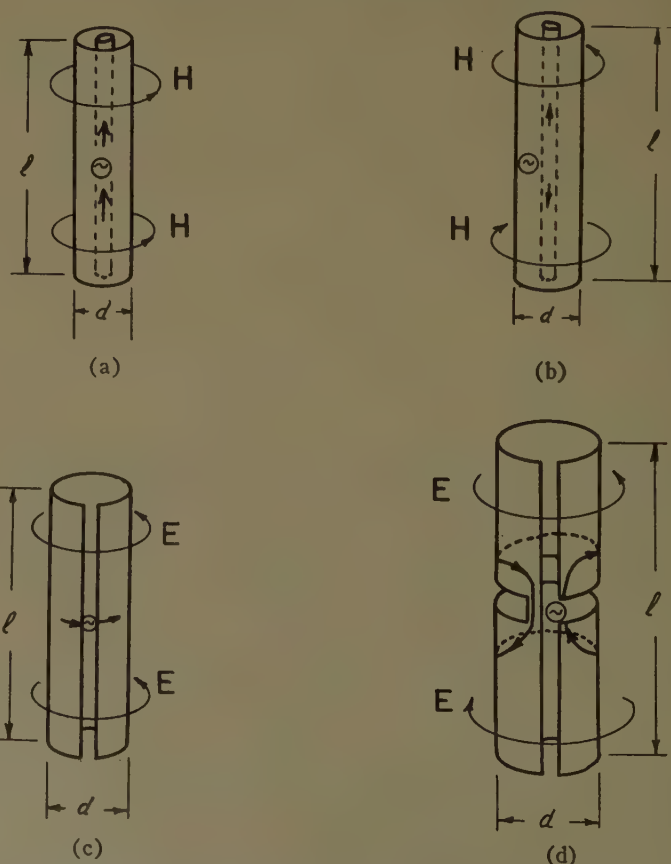


Fig. 5—Dipole and quadrupole antennas. (a) Electric dipole; (b) electric quadrupole; (c) magnetic dipole; (d) magnetic quadrupole, $l \gg d$.

VI. BOUNDARY CONDITIONS FOR NORMAL DIPOLE AND QUADRUPOLE SHEETS

Now that we see how a source can be replaced by normal dipoles or normal quadrupoles over a closed surface, it is of interest to work out the properties of such surface sources. The dipole case can be solved by first showing that a surface density U of normal electric dipoles on a closed surface is equivalent to a surface density

$$K = n \times \nabla \left(\frac{U}{Y} \right) \quad (40)$$

of tangential magnetic dipoles. Then it follows that tangential H , normal ZH and normal YE are continuous and tangential E is discontinuous, by the amount $-\nabla_T(U/Y)$ in the direction n . The boundary conditions for a surface density of normal magnetic dipoles are exactly analogous. Note that since $n \cdot YE$ and $n \cdot ZH$ are continuous, a normal dipole sheet, backed by a medium which makes $n \cdot E = 0$ and $n \cdot H = 0$, gives a field for which $n \cdot E = 0$ and $n \cdot H = 0$ on S . It follows from the theorem of Section IV that the field is zero everywhere. This is analogous to the result that a sheet of tangential electric dipoles, backed by a perfect conductor, gives zero field.

The sheet of normal dipoles has a simple connection with the Hertz potentials f_e and f_m introduced in (2) and (3). Obviously, normal electric dipoles do not react with the TE field generated by f_e . The effect on the TM field is represented by a discontinuity of $(-U)$ in $\partial f_m / \partial n$, f_m being continuous. However, the Hertz potential method is limited to certain coordinate systems [15] and to a uniform medium.

For a surface density F_q of normal electric quadrupoles, it can be shown by using (28) that

$$\langle qt \rangle = \oint E_t \cdot \left[Y h_1 \nabla_T \left(\frac{F_q}{Y h_1} \right) - n \frac{F_q}{Y} \nabla \cdot n Y \right] dS. \quad (41)$$

The last term represents the reaction between the test source t and a sheet of normal dipoles and can therefore be transformed by (40) to read

$$\langle qt \rangle = \oint \left[E_t \cdot Y h_1 \nabla_T \left(\frac{F_q}{Y h_1} \right) - H_t \cdot n \times \nabla \left(\frac{F_q \nabla \cdot n Y}{Y^2} \right) \right] dS. \quad (42)$$

This shows that the electric quadrupole density F_q is equivalent to tangential electric dipoles

$$J = h_1 Y \nabla_T \left(\frac{F_q}{h_1 Y} \right) \quad (43)$$

and tangential magnetic dipoles

$$K = n \times \nabla_T \left(\frac{F_q \nabla \cdot n Y}{Y^2} \right). \quad (44)$$

The tangential electric dipoles give a discontinuity in tangential H of amount

$$h_1 Y \nabla_T \left(\frac{F_q}{h_1 Y} \right) \times n, \quad (45)$$

and a discontinuity in normal YE of amount

$$-h_1 \nabla_T \cdot \left[Y \nabla_T \left(\frac{F_q}{h_1 Y} \right) \right], \quad (46)$$

as we saw in (37); tangential E and normal ZH are continuous. The tangential magnetic dipoles give a discontinuity in tangential E of amount

$$-\nabla_T \left(F_q \frac{\nabla \cdot n Y}{Y^2} \right) \quad (47)$$

but no discontinuity in $n \cdot ZH$ or tangential H or $n \cdot YE$. Thus, F_q gives a discontinuity (45) in tangential H , a discontinuity (46) in normal YE , a discontinuity (47) in tangential E , and continuity of normal ZH . The magnetic quadrupole sheet is analogous. In cases where the Hertz potential method can be used, the sheet of normal electric quadrupoles has no effect of f_e and gives a discontinuity F_q in f_m and a discontinuity $-F_q \nabla \cdot n$ in $\partial f_m / \partial n$.

The term $\nabla \cdot n Y$ in (44) has an interesting effect where the curvature of S is very large. For example, $\nabla \cdot n = 1/\rho$ for a cylinder of radius ρ . Then we see from (44), that $K \sim 1/\rho$ so $K \rightarrow \infty$ as $\rho \rightarrow 0$. This means that at an edge, the surface density K becomes a line density concentrated on the edge. By taking Y to be uniform, we find that the line density M (or "edge current" as it is sometimes called) equals $(A/Y) \nabla F_q$, where A is the wedge angle.

VII. A COMBINATION OF DIPOLES AND QUADRUPOLES WHICH GIVES ZERO ON ONE SIDE

We saw in the Introduction that the combination of J_h and K_h in (1) gave the primary field on one side of S and zero on the other side. This suggests that an appropriate combination of normal dipoles and quadrupoles might give the primary field on one side and zero on the other. To investigate this possibility we replace the normal dipoles and quadrupoles by the equivalent tangential dipoles J and K given in Section VI. Thus, if U and V are the surface densities of normal electric and magnetic dipoles, and F and G are the surface densities of normal electric and magnetic quadrupoles, then

$$J = \nabla_T \left(\frac{V}{Z} \right) \times n + h_1 Y \nabla_T \left(\frac{F}{h_1 Y} \right) + \nabla_T \left(\frac{G \nabla \cdot n Z}{Z^2} \right) \times n \quad (48)$$

and

$$K = n \times \nabla_T \left(\frac{U}{Y} \right) + h_1 Z \nabla_T \left(\frac{G}{h_1 Z} \right) + n \times \nabla_T \left(\frac{F \nabla \cdot n Y}{Y^2} \right). \quad (49)$$

We see that (48) and (49) can be written in the form

$$J = \nabla_T P \times n + Y h_1 \nabla_T Q, \quad (50)$$

$$K = n \times \nabla_T R + Z h_1 \nabla_T S, \quad (51)$$

where $PQRS$ are independent functions if $UVfg$ are independent. Now we know that if $J = n \times H_g$ and $K = E_g \times n$, as in (1), the proposition is proved. It therefore remains to show that we can always find $PQRS$ so that any J and K are represented by (50) and (51).

Clearly it is sufficient to consider J alone since K has the same form. From (50) after some manipulation, we obtain

$$\nabla_T \cdot \left(\frac{J}{h_1} \right) = \nabla_T \cdot Y \nabla_T Q \text{ on } S. \quad (52)$$

Let Q be a solution of (52). Let

$$J' = h_1 Y \nabla_T Q. \quad (53)$$

Then

$$\nabla_T \cdot \left(\frac{J - J'}{h_1} \right) = 0$$

whence it follows, after a few steps, that

$$J - J' = \nabla_T P \times n \quad (54)$$

where P is an arbitrary function. On substitution from (53) for J' , this gives J in the form (50). Hence, given any continuous vector function J defined on S , we can always find P and Q to fit the representation (50), and the proposition is thereby proved.

From (52) with $J = n \times H$, we get

$$\begin{aligned} \nabla_T \cdot Y \nabla_T Q &= \nabla_T \cdot \left(\frac{n \times H}{h_1} \right) \\ &= - \frac{n}{h_1} \cdot \nabla \times H = - n \cdot \frac{Y E}{h_1}, \end{aligned} \quad (55)$$

or since

$$Q = \frac{F}{Y h_1} \quad (56)$$

$$h_1 \nabla_T \cdot Y \nabla_T \left(\frac{F}{Y h_1} \right) = - Y E_n \quad (57)$$

which is the same as (37). Thus, $F = f$, as defined in Section V. (58)

The function P in (50) can be found by using the identity

$$\nabla_T \cdot \left(\frac{n}{h_1} \times \nabla \phi \right) = 0 \quad \text{for any } \phi. \quad (59)$$

Thus, from (50) putting $J = n \times H$,

$$\nabla_T \cdot \left(\frac{H}{Y h_1^2} \right) = - \nabla_T \cdot \left(\frac{\nabla_T P}{Y h_1^2} \right) \quad (60)$$

which determines P from tangential H . By (49) and (50), V and G are found from

$$P = \frac{V}{Z} + \frac{G \nabla \cdot n Z}{Z^2} = \frac{V}{Z} + \frac{g \nabla \cdot n Z}{Z^2} \quad (61)$$

and by (38) g is found from

$$h_1 \nabla_T \cdot Z \nabla_T \left(\frac{g}{h_1 Z} \right) = - Z H_n$$

Similarly, from (51) with $K = E \times n$,

$$\nabla_T \cdot \left(\frac{E}{Z h_1^2} \right) = - \nabla_T \cdot \left(\frac{\nabla_T R}{Z h_1^2} \right) \quad (62)$$

and

$$R = \frac{U}{Y} + f \frac{\nabla \cdot n Y}{Y^2}. \quad (63)$$

Thus, the combination of normal electric and magnetic dipoles U and V and normal electric and magnetic quadrupoles f and g gives the primary field on one side of S and zero field on the other, if f and g satisfy (37) and (38), and U and V satisfy (60), (61), (62), and (63).

This result simplifies considerably if S is plane, because $h_1 = 1$ and therefore

$$\nabla_T \cdot E = - \frac{\partial E_n}{\partial n}.$$

Hence,

$$\nabla_T^2 f = - Y E_n \quad \nabla_T^2 g = Z H_n \quad (64)$$

$$\nabla_T^2 U = - Y \frac{\partial E_n}{\partial n} = Y \nabla_T \cdot E$$

$$\nabla_T^2 V = - Z \frac{\partial H_n}{\partial n} = Z \nabla_T \cdot H. \quad (65)$$

Since the combination gives zero on one side of S , we can introduce a perfect conductor ($n \times E = 0$) on that side. This annuls the contributions from V and f and doubles the contributions from U and g . Thus, the Huygens source can be taken as the electric dipoles $2U$ and magnetic quadrupoles $2g$ in free space. As an example, consider diffraction through a hole in a plane metal screen. On the screen $H_n = 0$ and $\partial E_n / \partial n = - \nabla_T \cdot E = 0$, so by (64) and (65) we can take $U = 0$ and $g = 0$ on the screen. The diffracted field is therefore the same as the field of the electric dipoles $2U$ and magnetic quadrupoles $2g$ in the aperture location but with the screen removed.

From the previous paragraph we see that the Huygens source can be taken as the combination of $2U$ and $2g$ in free space. Now in Section V we found that the Huygens source could be taken as the combination of $2f$ and $2g$. It follows that U is equivalent to f and V is equivalent to g for a plane surface. Thus, the Huygens source can be taken alternatively as the combination of electric dipoles $2U$ and magnetic dipoles $2V$.

These results have some applications in the general problem of direction finding with an antenna consisting of an aperture in a horizontal metal surface. The direction of an incoming wave is to be found from the variation of received voltage, as the antenna is rotated around a vertical axis. One of the key problems in this type of measurement is known as polarization error: in essence the calibration depends drastically on the polarization of the wave unless the antenna is designed very carefully. Now if the antenna is equivalent to an array of vertically-polarized electric dipoles it will pick up only the vertically-polarized component of E , regardless of the direction of the incoming wave, as for example in the Adcock direction finder. It is thus free of polarization error. The condition for this is $V=0$, or from (65)

$$0 = \frac{\partial H_n}{\partial n} \quad \text{or} \quad 0 = \nabla_T \cdot H,$$

which is satisfied by any TM mode. Similarly, if the direction finder is to pick up only the horizontal E component, the aperture field must satisfy

$$0 = \frac{\partial E_n}{\partial n} \quad \text{or} \quad 0 = \nabla_T \cdot E$$

which is satisfied by any TE mode. For an arbitrary antenna the polarization error is the result of the combined contributions from the electric and magnetic vertical dipoles, which are given by the formulas in this section.

Returning to the surface of arbitrary shape, let us see what is the effect of an ideal metal backing ($n \times E = 0$) on the combination of electric and magnetic dipoles and quadrupoles denoted by UVf and g at the beginning of this section. As we have seen in Section II, the result is equivalent to tangential magnetic dipoles K on the metal surface, K being given in terms of $UVfg$ by (49). The first and third terms of (49) represent normal electric dipoles of density

$$U + \frac{f \nabla \cdot n Y}{Y}.$$

The second term represents normal magnetic quadrupoles of density g . Thus, there is a certain distribution of normal electric dipoles and magnetic quadrupoles on a metal surface of arbitrary shape which is equivalent to the primary source. In practice, it would take the form of an array of coaxial waveguides for the electric dipoles and circular waveguides for the magnetic quadrupoles as described in Section V.

VII. A MULTIPOLE COMBINATION OF VERTICAL DIPOLES WHICH IS EQUIVALENT TO A HORIZONTAL DIPOLE

One of the consequences of Section IV is that there exists a distribution of vertical dipoles on a horizontal plane which is equivalent to a horizontal dipole. This

suggests that there might be a combination of vertical dipoles, in multipole formation at the same point, which is equivalent to a horizontal dipole. To investigate this possibility we take the horizontal dipole in free space and pointing in the x direction, the z direction being vertical.

We wish to find a combination of electric and magnetic multipoles composed of dipoles pointing in the z direction, which is equivalent to an electric dipole pointing in the x direction. Since any z -directed electric dipole gives zero H_z , the contributions to H_z must come entirely from the magnetic multipoles. We therefore seek a magnetic multipole combination which gives the same H_z as the x -directed electric dipole. In the same way we adjust the electric multipole combination to reproduce E_z . Then it follows from the Hertz potential method, or from Section IV of this paper, that all components of E and H , for two different sources, are equal if E_z and H_z are equal.

Let \mathbf{c} represent the magnetic multipole combination and \mathbf{d} the electric dipole. If \mathbf{t} is a z -directed unit magnetic dipole,

$$\langle t\mathbf{d} \rangle = -H_{dz}(\text{at } t)$$

$$\langle t\mathbf{c} \rangle = -H_{cz}(\text{at } t).$$

Therefore, we require

$$\langle t\mathbf{d} \rangle = \langle t\mathbf{c} \rangle \quad (66)$$

for all positions of \mathbf{t} . Now

$$\langle t\mathbf{d} \rangle = \langle d\mathbf{t} \rangle = E_{tx}(\text{at } d) \quad (67)$$

assuming \mathbf{d} to have unit moment, and

$$\langle t\mathbf{c} \rangle = \langle c\mathbf{t} \rangle = aH_{zt} + b \frac{\partial H_{zt}}{\partial x} + c \frac{\partial H_{zt}}{\partial y} \dots = \left(a + b \frac{\partial}{\partial x} + \dots + e \frac{\partial^2}{\partial x^2} \dots + p \frac{\partial^{l+m+n}}{\partial x^l \partial y^m \partial z^n} \dots \right) H_{zt} \quad (68)$$

where $abc \dots e \dots p \dots$ are unknown coefficients: only H_{zt} occurs in (68) because we take \mathbf{c} to consist of z -directed magnetic dipoles in all possible multipole arrangements. Thus, the first term represents a dipole, and the second a quadrupole consisting of two opposing dipoles separated by an infinitesimal shift along the x axis. Since \mathbf{t} is a z -directed unit magnetic dipole,

$$E_{tx} = \frac{\partial \phi}{\partial y} \quad \text{where} \quad \phi = \frac{e^{-i\beta r}}{4\pi r}, \quad (\beta^2 = -YZ) \quad (69)$$

$$ZH_{tz} = -\frac{\partial^2 \phi}{\partial z^2} - \beta^2 \phi \quad (70)$$

Substitution from (69) and (70) in (67) and (68), combined with (66) gives

$$\frac{\partial \phi}{\partial y} = -\frac{1}{Z} \left(\frac{\partial^2}{\partial z^2} + \beta^2 \right) \left(a + b \frac{\partial}{\partial x} \dots \right) \phi. \quad (71)$$

This is to be true for all positions of \mathbf{t} , which means for all values of r in $\phi(r)$. By choosing the coefficients a , $b \dots$, we can put the right side of (71), which is $\langle ct \rangle$, into the form

$$\langle ct \rangle = \frac{\partial}{\partial y} \left(\frac{\partial^2}{\partial z^2} + \beta^2 \right) \left(a_0 \phi + a_2 \frac{\partial^2 \phi}{\partial z^2} + a_4 \frac{\partial^4 \phi}{\partial z^4} \dots \right). \quad (72)$$

Then (71) is satisfied for all ϕ , if

$$\phi = \beta^2 a_0 \phi + (a_0 + \beta^2 a_2) \frac{\partial^2 \phi}{\partial z^2} + (a_2 + \beta^2 a_4) \frac{\partial^4 \phi}{\partial z^4} \dots$$

or

$$\begin{aligned} a_0 &= \frac{1}{\beta^2} \\ a_2 &= -\frac{a_0}{\beta^2} = -\frac{1}{\beta^4} \\ a_4 &= -\frac{a_2}{\beta^2} = \frac{1}{\beta^6} \\ &\vdots \\ &\vdots \end{aligned} \quad (73)$$

From (70), (72) and (73), recalling that $\beta^2 = -YZ$,

$$\langle ct \rangle = \frac{1}{Y} \left(1 - \frac{1}{\beta^2} \frac{\partial^2}{\partial z^2} + \frac{1}{\beta^4} \frac{\partial^4}{\partial z^4} \dots \right) \frac{\partial H_{tz}}{\partial y}. \quad (74)$$

We now define this to be the reaction $\langle ct \rangle$ with an arbitrary test source \mathbf{t} , thus specifying \mathbf{c} uniquely. The first term represents a quadrupole consisting of two opposing z magnetic dipoles separated in the y direction: the next term represents a quadru-quadrupole and so on. Hence, the source \mathbf{c} so defined gives zero E_z and the same H_z as a unit x -directed electric dipole \mathbf{d} at the same point.

It remains to find another multipole combination which gives zero H_z and the same E_z as the unit x directed electric dipole \mathbf{d} . We therefore take \mathbf{t} as a unit z -directed electric dipole and make

$$\langle dt \rangle = \langle bt \rangle \quad (75)$$

for all positions of \mathbf{t} , where \mathbf{b} represents the desired multipole combination. In this case,

$$\begin{aligned} \langle dt \rangle &= E_{tz} \text{ (at } d) \\ &= \frac{1}{Y} \frac{\partial^2 \phi}{\partial x \partial z} \\ \langle bt \rangle &= \sum C_{lmn} \frac{\partial^{l+m+n} E_{zt} \text{ (at } d)}{\partial x^l \partial y^m \partial z^n} \end{aligned} \quad (76)$$

(taking \mathbf{b} as a combination of z -directed electric dipoles because this gives zero H_{zb}) and

$$E_{zt} = \frac{1}{Y} \left[\frac{\partial^2}{\partial z^2} + \beta^2 \right] \phi. \quad (77)$$

Thus, (75) gives

$$\frac{\partial^2 \phi}{\partial x \partial z} = \sum C_{lmn} \frac{\partial^{l+m+n}}{\partial x^l \partial y^m \partial z^n} \left[\frac{\partial^2 \phi}{\partial z^2} + \beta^2 \phi \right] \quad (78)$$

Hence, we choose C_{lmn} so that (76) reads

$$\langle bt \rangle = \sum_{N=0,2,4,\dots} C_N \frac{\partial^{N+2} E_{zt}}{\partial z^{N+1} \partial x}. \quad (79)$$

Then (78) is satisfied if

$$\phi = \sum_{N=0,2,4,\dots} C_N \frac{\partial^N}{\partial z^N} \left[\frac{\partial^2 \phi}{\partial z^2} + \beta^2 \phi \right]$$

which gives $C_N = (-1)^{N/2} \beta^{-N-2}$ as in (73).

We now define the multipole combination \mathbf{b} by (79) for an arbitrary test source \mathbf{t} . Thus,

$$\langle bt \rangle = \frac{1}{\beta^2} \left[1 - \frac{1}{\beta^2} \frac{\partial^2}{\partial z^2} + \frac{1}{\beta^4} \frac{\partial^4}{\partial z^4} \dots \right] \frac{\partial^2 E_{zt}}{\partial z \partial x}. \quad (80)$$

The first term represents an octupole consisting of z -directed electric dipoles with a shift in the z direction and a shift in the x direction, and so on for the other terms.

Finally, since the combination of \mathbf{c} and \mathbf{b} gives the same E_z and H_z as \mathbf{d} , it follows that for any test source \mathbf{t} ,

$$\langle dt \rangle = \langle ct \rangle + \langle bt \rangle;$$

$\langle ct \rangle$ and $\langle bt \rangle$ being specified by (74) and (80). We have thus represented a horizontal dipole as a combination of vertical (electric and magnetic) dipoles concentrated at the same point.

This result enables us to solve the problem of an arbitrary source above a plane on which $\mathbf{n} \cdot \mathbf{E} = 0$ and $\mathbf{n} \cdot \mathbf{H} = 0$, a problem which was raised in Section III. The method of solution is as follows. The source can be split into vertical and horizontal dipoles and, by the results of this section, the horizontal dipoles can be replaced by vertical dipoles in various multipole arrangements. Now the method of images can be applied to the vertical dipoles (it cannot be applied to horizontal dipoles because of the peculiar boundary conditions). It will be found that the image of a vertical dipole is equal and opposite to the object. The image of a vertical vertical quadrupole is the same as the object but for a vertical horizontal quadrupole it is opposite. In general, it will be found that a horizontal displacement of a vertical dipole has no effect on the phase of the image and a vertical displacement introduces a phase reversal. Comparing (74) and (80), we see that (80) has one more vertical displacement; i.e., an extra differentiation in the z direction. Thus, the image of the equivalent electric multipoles \mathbf{b} has one more phase reversal than the image of the equivalent magnetic multipoles \mathbf{c} , assuming the primary source consists of electric dipoles. (If their phases were the same, these images could obviously be

combined to give a horizontal dipole image.) Specifically, the image of a unit horizontal electric dipole is represented by $\mathbf{b}-\mathbf{c}$, where \mathbf{b} and \mathbf{c} are the sources defined in (80) and (74).

In this way, we obtain an explicit solution to the problem of an arbitrary source above a plane on which $\mathbf{E} \cdot \mathbf{n} = 0$ and $\mathbf{H} \cdot \mathbf{n} = 0$. The existence of such a solution, which satisfies Maxwell's equations and the boundary conditions, demonstrates that the boundary conditions are consistent with Maxwell's equations. This point was covered in Section III by a completely different approach, an approach which, however, left some question as to whether the solution for an arbitrary source was unique, because a plane wave at normal incidence is unaffected by these boundary conditions. Since we have found that the reflection is perfect (*i.e.*, the images are all of magnitude equal to the object), this fact does not affect the result for a source of finite size.

APPENDIX

THE BOUNDARY CONDITIONS $E_n = 0$ $H_n = 0$ APPLIED TO WAVEGUIDE MODES

Suppose that a waveguide is blocked by a uniform medium which extends over $z \geq 0$, the region $z \leq 0$ being filled with air. Let the medium be anisotropic as described in Section III so that the environment parameters Y and Z are represented by tensors. In rectangular coordinates, with the z axis as the waveguide axis, we take Y and Z to be diagonal; *i.e.*,

$$\begin{aligned} Z_{11} = Z_{22} = a & \quad Z_{33} = p \\ Y_{11} = Y_{22} = b & \quad Y_{33} = q. \end{aligned} \quad (81)$$

As we saw in Section III, $E_n = 0$ and $H_n = 0$ at the boundary of this medium when $p = 0$ and $q = 0$. However, in order to understand the properties of this peculiar medium, we start with finite values of p and q and then see what happens as they approach zero. The problem is to find the field due to a wave traveling down the air filled portion ($z \leq 0$) towards the boundary at $z = 0$.

As a preliminary, let us see whether the Hertz potential method works in the anisotropic medium. We try a solution in the form

$$\mathbf{E} = \nabla \times \hat{\mathbf{z}} f. \quad (82)$$

Then from Maxwell's equation,

$$-ZH = \nabla \times \mathbf{E} = \nabla \times \nabla \times \hat{\mathbf{z}} f = \nabla \frac{\partial f}{\partial z} - \hat{\mathbf{z}} \nabla^2 f.$$

Assuming an exponential variation with z of the form $\exp \gamma z$, we have, on using (81),

$$aH_x = -\gamma \frac{\partial f}{\partial x}, \quad aH_y = -\gamma \frac{\partial f}{\partial y}, \quad (83)$$

$$pH_z = \nabla^2 f - \gamma^2 f. \quad (84)$$

The remaining Maxwell equation

$$\mathbf{Y} \mathbf{E} = \nabla \times \mathbf{H}$$

gives, on substitution for \mathbf{E} , \mathbf{H} and \mathbf{Y} ,

$$\begin{aligned} b \frac{\partial f}{\partial y} &= \frac{1}{p} (\nabla^2 - \gamma^2) \frac{\partial f}{\partial y} + \frac{\gamma^2}{a} \frac{\partial f}{\partial y}, \\ b \frac{\partial f}{\partial x} &= \frac{1}{p} (\nabla^2 - \gamma^2) \frac{\partial f}{\partial x} + \frac{\gamma^2}{a} \frac{\partial f}{\partial x}, \\ 0 &= \frac{-\gamma}{a} \frac{\partial^2 f}{\partial x \partial y} + \frac{\gamma}{a} \frac{\partial^2 f}{\partial x \partial y}. \end{aligned}$$

These equations are satisfied if

$$\left(\frac{\nabla^2}{p} - \frac{\gamma^2}{p} + \frac{\gamma^2}{a} - b \right) f = 0.$$

Putting $f = e^{\gamma z} F(xy)$, we have

$$\nabla^2 F + k^2 F = 0 \quad (85)$$

where

$$k^2 = p \frac{\gamma^2}{a} - pb. \quad (86)$$

It follows that the Hertz potential method is valid. The mode numbers k are fixed by the boundary conditions at the walls of the waveguide. For a rectangular waveguide of height A and width B ,

$$k^2 = \left(\frac{L\pi}{A} \right)^2 + \left(\frac{M\pi}{B} \right)^2 \quad \text{for } L \text{ or } M = 0, 1, 2, 3, \dots$$

Since k is thereby fixed, it follows from (86) that

$$\gamma \rightarrow -\infty \quad \text{as } p \rightarrow 0,$$

and if p and a are pure imaginary, as in a lossless medium, γ is pure real. Thus, as $p \rightarrow 0$, all modes die out exponentially with infinite rapidity, even though there is no loss. The effect of letting $p \rightarrow 0$ can be pictured as similar to the effect of introducing perfectly conducting filaments parallel to the z axis and very close together. This is like subdividing the waveguide into many small waveguides each of which would have an extremely high cutoff frequency. In the limit, any frequency would be cut off, resulting in the exponential decay of all modes.

Suppose that the incident wave is a TE mode so that it can also be represented by a Hertz potential function. Since the tangential components of \mathbf{E} and \mathbf{H} must be continuous at $z = 0$, the transverse function F must be the same for the incident, reflected, and transmitted waves. Thus, we can write for the Hertz potentials f ,

$$z \leq 0 \quad f_0 = (Ae^{\gamma_0 z} + Be^{-\gamma_0 z}) F(xy), \quad (87)$$

$$z \geq 0 \quad f = Ce^{\gamma z} F(xy). \quad (88)$$

Tangential \mathbf{E} and \mathbf{H} are continuous if f and $(1/a)(\partial f / \partial z)$ are continuous, where a is defined in (81).

Hence,

$$A + B = C,$$

$$\frac{\gamma_0}{a_0}(A - B) = \frac{\gamma C}{a}.$$

Therefore,

$$B = A \frac{\gamma_0' - \gamma'}{\gamma_0' + \gamma'}, \quad (89)$$

$$C = A \frac{2\gamma_0'}{\gamma_0' + \gamma'}, \quad (90)$$

$$a\gamma' = \gamma \quad a_0\gamma_0' = \gamma_0. \quad (91)$$

Let us take $a = a_0$, since it is the effect of p that is of interest. Then since $\gamma \rightarrow -\infty$ as $p \rightarrow 0$,

$$B \rightarrow -A \quad (92)$$

$$C \rightarrow 2A \frac{\gamma_0}{\gamma} \rightarrow 0. \quad (93)$$

These equations show that $E_x = 0$ and $E_y = 0$ at the boundary, which, therefore, looks like a short circuit to a TE mode. For $z \geq 0$, we find E_x and $E_y = 0$; H_x and H_y are finite, but H_z is infinite. From (84), (85), and (86)

$$pH_z = -k^2CF(xy)e^{\gamma z} \quad p \rightarrow 0, \gamma \rightarrow -\infty. \quad (94)$$

This singularity in H_z corresponds to a sheet of normal magnetic dipoles. For $p \neq 0$, the volume density of these magnetic dipoles is finite. It can be found by comparing the equations

$$\nabla \times E = -ZH \quad \nabla \times H = YE \quad (95)$$

and

$$\nabla \times E = -K_i - Z_0H \quad \nabla \times H = Y_0E + J_i. \quad (96)$$

The first pair applies to the medium characterized by Y and Z , and the second pair applies to air, characterized by Y_0 and Z_0 . The field, represented by E or H , is the same in both cases. It follows that

$$K_i = (Z - Z_0)H, \quad J_i = (Y - Y_0)E. \quad (97)$$

The interpretation of (96) is that the combination of the primary source and the magnetic and electric dipoles K_i and J_i in air is equivalent to the primary source in the presence of the medium YZ . Thus, K_i is the volume density of current moment of the induced magnetic dipoles. From (97),

$$K_{iz} = (p - p_0)H_z \rightarrow -p_0H_z \\ \rightarrow \frac{p_0k^2C}{p} F(xy)e^{\gamma z} \quad \text{from (94)}$$

$$\rightarrow \frac{p_0k^2}{p} 2A \frac{\gamma_0}{\gamma} F(xy)e^{\gamma z} \quad \text{from (93)}$$

$$\rightarrow p_0 \frac{\gamma^2}{a_0} 2A \frac{\gamma_0}{\gamma} F(xy)e^{\gamma z} \quad \text{from (96)}.$$

Therefore,

$$\int_0^\infty K_{iz} dz \rightarrow -2A\gamma_0 F(xy) \quad \text{since } p_0 = a_0. \quad (98)$$

The induced source is thus equivalent to a surface density $-2A\gamma_0 F(xy)$ of normal magnetic dipoles. For an incident TM mode, the result is obviously the analogous distribution of normal electric dipoles. For an arbitrary incident field, we combine the TE and TM parts and thus we get an explicit formula for the functions ϕ_b and ψ_b in Section IV. Since we have found that the field beyond the surface is zero, we see that this analysis checks the fact that the field of the sheet of normal electric and magnetic dipoles is equal and opposite to the incident field, a fact which we discovered in Section IV to be true for a surface of arbitrary shape.

When the cross section of the waveguide tends to infinity, the mode expansion goes over into a Fourier integral representation, and thus becomes a general representation for the field of an arbitrary source in front of a plane boundary. The spectrum now includes the exceptional case of normal incidence, but this component is an infinitesimal fraction of the whole spectrum (for any finite source). This provides an alternative explanation of why the case of normal incidence does not upset the results of Section IV.

REFERENCES

- [1] J. Larmor, *Proc. London Math. Soc.*, ser. 2, 1, 1; January, 1903.
- [2] A. E. H. Love, *Proc. London Math. Soc.*, ser. 2, 1, 37; January, 1903.
- [3] H. M. McDonald, *Proc. London Math. Soc.*, ser. 10, 91, 1911 and *Phil. Trans. (A)* 212, 295; 1912.
- [4] S. A. Schelkunoff, *Bell Sys. Tech. J.*, vol. 15, p. 92; 1936.
- [5] J. A. Stratton, and L. J. Dhu, *Phys. Rev.*, vol. 56, p. 99; 1939.
- [6] J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., Inc., New York, N. Y., p. 467; 1941.
- [7] S. A. Schelkunoff, "Electromagnetic Waves," D. VanNostrand Co., Inc., New York, N. Y., p. 121; 1943.
- [8] B. B. Baker and E. T. Copson, "The Mathematical Theory of Huygens' Principle," Clarendon Press, Oxford, England, 2nd ed.; 1950.
- [9] V. H. Rumsey, *Phys. Rev.*, vol. 94, pp. 1483-1491; June, 1954.
- [10] E. T. Whittaker, *Proc. London Math. Soc.*, vol. 1, p. 367; 1903. Explicit formulas for f_s and f_m .
- [11] P. Debye, *Ann. Phys.*, vol. 30, p. 57; 1909.
- [12] T. J. I' A. Bromwich, *Phil. Trans.*, A220, 175; 1920.
- [13] S. A. Schelkunoff, "Advanced Antenna Theory," John Wiley and Sons, Inc., New York, N. Y., p. 11; 1952. Simple proof of generality of the method in spherical coordinates.
- [14] H. S. Green and E. Wolf, *Proc. Phys. Soc.*, A66, 1129; 1953.
- [15] P. M. Morse and H. Feshbach, "Methods of Theoretical Physics," McGraw-Hill Book Co., Inc., New York, N. Y., p. 1764; 1953.
- [16] C. J. Bouwkamp and H. B. G. Casimir, *Physica*, 20, 539; 1954.
- [17] A. Nisbet, *Proc. Roy. Soc.*, ser. A231, 250; August, 1955.
- [18] V. H. Rumsey, The Ohio State University Research Foundation Rept. No. 444-15; January, 1954.
- [19] V. H. Rumsey, "Huygens' Principle as an Exact Physical Concept," Electronics Res. Lab., University of California, Berkeley, Rept., Nov., 1954.

A Solution to the Equiangular Spiral Antenna Problem*

V. H. RUMSEY†

ABSTRACT

IN ORDER to simplify the mathematical problem and yet retain the essential features of the frequency-independent modes,^{1,2} we consider an antenna consisting of an infinite number of conducting filaments, uniformly spread around a cone $\theta = \theta_0$, each filament being a spiral of the form $r = \exp(-a\phi)$, r , θ and ϕ being conventional spherical coordinates and a a real constant. From the symmetry of the problem it is apparent that we can consider solutions which vary with ϕ as $\exp(jn\phi)$ where $j^2 = -1$ and n is an integer. The case $n=1$ is approximately equivalent to the balanced excitation of a practical two-conductor self-complementary antenna.^{1,2} For the sake of illustration, consider the case of the plane structure ($\theta_0 = \pi/2$). From the symmetry with respect to this plane it will be found that the boundary conditions for the tangential electric field E are the same (apart from a constant C) as for the tangential magnetic field H at all points on the plane as it is approached from one side. On passing through the plane, tangential E must be continuous and tangential H discontinuous. Therefore, the constant C has the opposite sign on the other side. Since the boundary conditions on tangential E (or H) determine the field uniquely, we therefore consider solutions of Maxwell's equations for which

$$E = CH \quad (1)$$

E and H being complex vector functions of position.³ It is found immediately that such solutions exist if

$$C = \pm j\eta H \quad \eta \simeq 377 \text{ ohms.} \quad (2)$$

They can be expressed in terms of a single scalar in several ways, of which the following has proved most useful.

$$E_1 = -\beta \nabla \times \hat{z} U_1 + \nabla \times \nabla \times \hat{z} U_1 \quad z \geq 0 \quad (3)$$

$$E_2 = +\beta \nabla \times \hat{z} U_2 + \nabla \times \nabla \times \hat{z} U_2 \quad z \leq 0 \quad (4)$$

where E_1 and E_2 denote the fields on either side of the

antenna and U_1 and U_2 are solutions of

$$\nabla^2 U + \beta^2 U = 0. \quad (5)$$

Tangential E is continuous and tangential H discontinuous at $z=0$ if

$$U_1 = -U_2 \quad \text{and} \quad \frac{\partial U_1}{\partial z} = \frac{\partial U_2}{\partial z} \quad \text{at } z = 0, \quad (6)$$

which means that U_2 is the negative reflection of U_1 in $z=0$. Tangential E parallel to the filaments vanishes and the current (represented by the discontinuity in tangential H) flows along the filaments if

$$a \left(\frac{-\beta}{\rho} jn U_1 + \frac{\partial^2 U_1}{\partial \rho \partial z} \right) = \frac{jn}{\rho} \frac{\partial U_1}{\partial z} + \beta \frac{\partial U_1}{\partial \rho} \quad \text{at } z = 0 \quad (7)$$

for all $0 \leq \rho \leq \infty$; ρ , ϕ , and z being the cylindrical coordinates. This equation combined with the boundary conditions at $r=0$ and $r=\infty$ completes the specification of the problem in terms of U_1 . Eq. (7) can be solved by taking the Fourier Bessel transform. The result obtained by B. Cheo is expressed by

$$U_1 = e^{jn\phi} \int_0^\infty f(y) e^{-j\beta z \sqrt{1-y^2}} J_n(\beta \rho y) y dy \quad (8)$$

$$f(y) = \left[\frac{1 - \sqrt{1-y^2}}{1 + \sqrt{1-y^2}} \right]^{n/2} \frac{(1 + ja\sqrt{1-y^2})^{-1-j(n/a)}}{y^2} \quad (9)$$

where J_n is the Bessel function of order n . The radiation pattern can be calculated from the transform by using the method of stationary phase. The result obtained by W. J. Welch is

$$E_\phi = \frac{e^{jn\phi} \tan^n(\theta/2)}{(\tan \theta) (1 + ja \cos \theta)^{1+j(n/a)}} \quad (10)$$

It follows directly from (1) that the pattern is circularly polarized for all θ and ϕ . The boundary condition that tangential E parallel to the filaments be zero ensures that the pattern goes to zero in the plane of the antenna. For $n=1$, (10) shows that the beam sharpens with decreasing a with a maximum along the z axis, $\theta=0$. For $n>1$, (10) gives a conical beam which sharpens with increasing n or decreasing a .

The current distribution is being worked out on a computer. It is known exactly for points near the input terminals from the static solution, which gives a uniform line density of current traveling along the filament with the velocity of light.

* Sponsored by the U. S. Army Signal Corps under a contract with the University of California, Berkeley, Calif.

† University of California, Berkeley, Calif.

¹ V. H. Rumsey, "Frequency independent antennas," IRE CONVENTION RECORD; 1957.

² J. D. Dyson, "The equiangular spiral antenna," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-7, pp. 181-187; April, 1959.

³ V. H. Rumsey, "General Antenna Analysis with Specific Reference to Frequency Independent Antennas," University of California, Berkeley, Calif., Electronics Res. Lab. Rept.; April, 1959.

General Theorems on the Transmission Coefficient from a Transmitting to a Receiving System

JEAN ROBIEUX†

Summary—Some theorems on the transmission coefficient from a transmitting to a receiving system will be presented. They can be considered as very general theorems in physics, since they can be applied to fields other than electromagnetic theory, such as wave mechanics and fluid mechanics. In electromagnetics they have been applied to the theory of surface waves and to propagation problems, particularly tropospheric scatter propagation.

LET $\bar{E}_1\bar{H}_1$ be the complex electromagnetic field existing when unit power normalized wave is radiated by transmitter antenna A . Let $\bar{E}_2\bar{H}_2$ be the complex field existing when unit power is radiated by antenna B . The receiver antenna is linked to the detector by a transmission line matched at each end. The field in this receiver transmission line when unit power is radiated by A is the field of normalized wave propagating from the antenna to the detector multiplied by a complex number $\bar{T}_{12} = \bar{T}_{12}e^{i\theta}$, which is the transmission coefficient from A to B . T_{12}^2 is the received power when a normalized wave is transmitted, and θ is the phase of the received wave. In the same way the transmission coefficient from B to A can be defined.

If S_2 is a surface immediately close to the receiver

$$\bar{T}_{12} = \frac{1}{4} \int_{s_2} (\bar{E}_1 \times \bar{H}_2 + \bar{H}_1 \times \bar{E}_2) \cdot d\mathbf{s}. \quad (1)$$

If S_1 is a surface close to the transmitter

$$\bar{T}_{21} = \frac{1}{4} \int_{s_1} (\bar{E}_2 \times \bar{H}_1 + \bar{H}_2 \times \bar{E}_1) \cdot d\mathbf{s}. \quad (2)$$

If $\bar{\epsilon}$ and $\bar{\omega}$ are symmetric

$$\nabla \cdot (\bar{E}_1 \times \bar{H}_2 + \bar{H}_1 \times \bar{E}_2) = 0. \quad (3)$$

Therefore, in this case,

$$\bar{T}_{12} = \bar{T}_{21} = \frac{1}{4} \int_s (\bar{E}_1 \times \bar{H}_2 + \bar{H}_1 \times \bar{E}_2) \cdot d\mathbf{s}, \quad (4)$$

S being any surface surrounding B completely without surrounding A .

The preceding relationships are of great interest because the medium is not necessarily homogeneous as it is in the Kirchhoff and Kottler relationships. Eqs. (1) and (2) assume no restriction on the medium between A and B . Eq. (4) allows heterogeneity of any kind if $\bar{\epsilon}$ and $\bar{\omega}$ are symmetrical. Huygen's principle is a particular case of these very general and rigorous relationships.

Two important cases are encountered for which $\bar{\epsilon}$ and $\bar{\omega}$ are not symmetrical; therefore, $\bar{T}_{12} \neq \bar{T}_{21}$.

In electron tubes it is found that

$$\bar{T}_{12} - \bar{T}_{21} = \int_v (\bar{E}_1 \bar{i}_2 - \bar{E}_2 \bar{i}_1) dv,$$

V being the volume between A and B , and \bar{i}_1 and \bar{i}_2 being the currents produced by normalized waves transmitted by A and B .

In the ferrite case,

$$\bar{T}_{12} - \bar{T}_{21} = \int_v \bar{H}_0 \cdot (\bar{H}_1 \times \bar{H}_2) dv,$$

\bar{H}_0 being the unit vector in the direction of the applied magnetic field. This last relationship has also been found by Heller at the Lincoln Laboratory, Massachusetts Institute of Technology.

Relationships giving the variation of \bar{T} with respect to frequency can be found. If $\bar{\epsilon}$ and $\bar{\omega}$ are symmetrical, this relationship has a simple form:

$$\frac{\delta \bar{T}}{\delta \omega} = \frac{1}{2} \int_v (\bar{E}_1 \cdot \bar{\epsilon} \cdot \bar{E}_2 - \bar{H}_1 \cdot \bar{\omega} \cdot \bar{H}_2) dv.$$

Using (4), I made the first demonstration in the early months of 1956, in the case of a surface wave problem.

These theorems can be considered as very general ones in physics. They can be extended to fields other than electromagnetic theory, such as wave mechanics and fluid mechanics. In electromagnetics I have used them to develop a theory of surface wave and propagation problems, particularly concerning tropospheric scatter propagation. A synthesis of all the work will be published in the *Annales de Radioélectricité*.

† Campagnie Generale de Telegraphie Sans Fil, Paris, France.

On Helmholtz's Theorem in Finite Regions*

J. VAN BLADEL†

DIGEST OF PAPER

THERE has been some discussion recently as to what constitutes a *complete* set of eigenvectors for a cavity, much of the discussion being concerned with multiply-bounded and multiply-connected regions. In the present paper, the completeness of the sets which have been commonly used in the literature is proved by examining the transformation

$$\nabla^2 \vec{f} = \vec{a}$$

with \vec{f} perpendicular to the boundary surface, and $\text{div } \vec{f} = 0$ thereon. The spectrum, and the eigenvectors, of the transformation are identical with those of an integral operator, and the completeness is proved for that operator. The eigenvectors are shown to be of two types: solenoidal and irrotational. The expansion coefficients

of any vector \vec{a} (not necessarily an electromagnetic field) are given. The irrotational terms form a term $\text{grad } \psi$, perpendicular to the boundary, and the solenoidal terms a term $\text{curl } \vec{v}$. The irrotational term is found to be zero for a solenoidal vector provided, in the case of a doubly-bounded volume, its flux through each individual surface is zero. The solenoidal term is zero for a vector \vec{a} with zero curl, \vec{a} being, in addition, perpendicular to the boundary.

Similar conclusions are reached about the completeness of the eigenvectors relative to the boundary conditions: \vec{f} tangent to the boundary surface, $\text{curl } \vec{f}$ perpendicular to the latter. Irrotational and solenoidal eigenvectors group each other, in the expansion, in terms $\text{grad } \theta$ and $\text{curl } \vec{w}$, where \vec{w} is perpendicular to the boundary. The term $\text{curl } \vec{w}$ is absent when the vector is irrotational. The term $\text{grad } \theta$ is not automatically zero when the vector is solenoidal. Many of the quoted results have been mentioned in the past. The purpose of the paper was to prove them in a rigorous fashion.

* The full paper is available as a MURA Rept. No. 440; December, 1958.

† Midwestern Universities Research Assn., and Dept. Elec. Engrg., University of Wisconsin, Madison, Wis.

The Synthesis of Large Radio Telescopes by the Use of Radio Interferometers

M. RYLE,[†] A. HEWISH,[†] AND J. R. SHAKESHAFT[†]

Summary—An outline is given of a method of simulating the resolving power of large radio telescopes by a process of synthesis, in which use is made of measurements taken with smaller structures arranged in different configurations. The method has been applied to the construction of some large radio telescopes at Cambridge.

THE OUTPUT of a large radio telescope, such as a broadside array, may be considered as the vector sum of the currents induced by an incident field in each of the individual elements of the array. In the case of a paraboloidal reflector the summation is effected at the focal point. When the direction of an incident wave does not coincide with the normal to the aperture plane, the induced currents in the elements of the aperture have a progressive phase shift which gives rise to the usual directional properties of the aperture. By adding the currents with their phases suitably adjusted it is possible to alter the direction of principal response of the aperture. This feature can be of great importance in systems where physical tilting of the array is impossible.

If one could measure, separately, the currents induced in each part of the aperture by moving a small antenna across it, then vector addition of the currents would give exactly the same result as that obtained by the use of the complete aperture. It is not possible to measure the phase with a single small antenna, but by using two small antennas connected as an interferometer the relative phases may be determined and the addition performed. Vector summation carried out in this way is the basis of a method of producing large apertures which may be called "aperture synthesis."

An alternative way of considering the principle follows from the fact that a simple interferometer, consisting of a pair of nondirectional receiving elements connected to a receiver, gives a response proportional to one Fourier component of the two-dimensional distribution of radio brightness over the sky. By taking a series of measurements in which the spacing and orientation of the interferometer are varied it is possible to determine the two-dimensional transform and hence, by Fourier inversion, to derive the brightness distribution. The resolving power increases as the separation of the antennas increases, or in other words, as the synthesized aperture increases.

Fig. 1 represents an aperture sub-divided into a number of small elements. If the current induced in the n th element by an incident plane wave is written as $I_n e^{i\phi_n}$ where ϕ_n is the phase of the wave at that element,

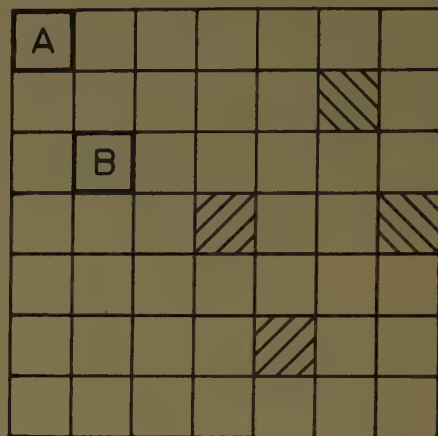


Fig. 1—The division of a large aperture into pairs of small elements, many of which are equivalent.

then the total power output P is given by

$$P \propto \sum_N I_n e^{i\phi_n} \cdot \sum_N I_n e^{-i\phi_n} \\ \propto \sum_N I_n^2 + \sum_N I_m \cdot I_n \cos(\phi_m - \phi_n).$$

If the N elements are all of the same size the first term is simply N times the power derived from a single element, and the other terms are proportional to the output of a phase-switching receiver [4] connected to elements m and n through equal cables. A measurement of the power from a single element, suitably combined with the outputs from a phase-switching receiver connected to a pair of elements which are arranged successively to cover all possible combinations of positions, then gives a result exactly equivalent to that obtained by using the whole large aperture. If the summation is carried out with the phases properly adjusted the direction of maximum response may be varied as if the aperture plane had been tilted. It is possible, without taking further observations, to scan the synthesized aperture through an angle limited only by the beamwidth of the small elements. In order to do this it is necessary to measure also the terms $I_m I_n \sin(\phi_m - \phi_n)$ by connecting the elements in phase quadrature to a separate receiver.

It is clear that many of the arrangements of the elements A and B are equivalent and need not be repeated. A procedure for obtaining the different terms without repetition is shown in Fig. 2, where one element, A , is kept fixed and the other, B , is moved over a rectangle of approximately twice the area. This enables the total number of observations to be reduced from N^2 , where N is the number of possible positions of a small element in the original aperture, to approximately $2N$.

[†] Mullard Radio Astronomy Observatory, Cavendish Laboratory, Cambridge, England.

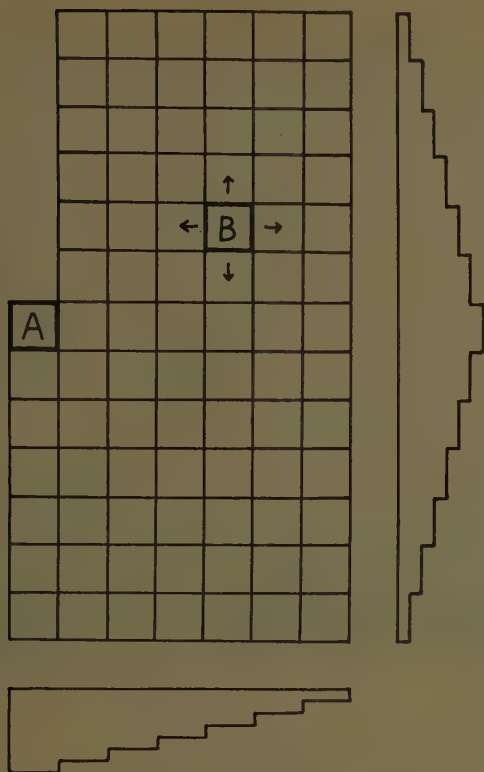


Fig. 2—By keeping element *A* fixed and by moving element *B* to occupy each of the small squares in turn all combinations of spacing and orientation possible in Fig. 1 are obtained. To duplicate the totality of arrangements in Fig. 1 the weighting factors shown at the right and below must be used to multiply the observations made when *B* is in the respective rows and columns.

To derive results equivalent to using the full aperture, the different terms must be added with a weighting factor appropriate to the number of repetitions of each configuration of the elements. The form of the weighting function appropriate to each position of the element *B* is also indicated in Fig. 2. Where synthesis is performed with no repetition of the arrangements there is no loss of information but the sensitivity is reduced. This may be no disadvantage at low frequencies when the signal strengths of radio sources are comparatively high.

In practice it may be convenient for the elements *A* and *B* to have different shapes, and for one of them to extend over the complete width of the required aperture so that the synthesis is performed in one dimension only. A system of this type is shown in Fig. 3. The element *A* must then have a tapered distribution of excitation along its length. It will be seen that this arrangement, when synthesized, is closely similar to the Mills Cross antenna [3]. Advantages of the one-dimensional synthesis are that there is only one antenna moving in one dimension, the time of observation is reduced and the computation (which it is convenient to carry out with an electronic machine) is less complicated.

It is necessary to consider how the synthesized aperture compares with a conventional aperture in regard to observation time and signal to noise ratio. Since a synthesis requires many observations it might be thought that the observing time is much longer. The extra time

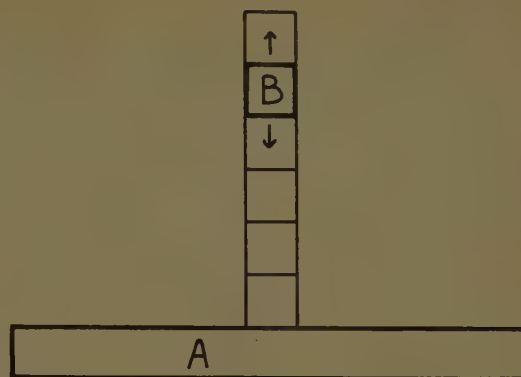


Fig. 3—Method of synthesis in one dimension using elements *A* and *B* of different shapes.

required for repeated observations is, however, compensated for by the fact that the beamwidth of the small elements is larger than that of the complete aperture, so that more sky is being observed at any one time.

It has been shown [2] that when a square aperture side *D* is used to survey the sky, information is lost if the angular intervals at which the beam is directed are greater than $\lambda/2D$, where λ is the wavelength.

The total observing time required to scan a solid angle Ω with this aperture is then $4D^2/\lambda^2 \cdot \Omega\tau$ where τ is the integration time constant of the receiver. If the same region of sky is scanned with the equivalent synthesized aperture obtained by using small elements of side *d*, the time required for each of the $2D^2/d^2$ arrangements is $4d^2/\lambda^2 \cdot \Omega\tau$. Thus the total observing time is approximately $8D^2/\lambda^2 \cdot \Omega\tau$ (for the same receiver time constant), which is only twice that required if the complete aperture were used. If the synthesis is performed in one dimension only, the observing time is no more than that time required when the complete aperture is used.

In both of these cases, if none of the observations is repeated, the signal to noise ratio is decreased by a factor approximately equal to the square root of the number of different arrangements used in the synthesis.

It should be noted that these arguments are only applicable when the area of sky surveyed is larger than the beam of the small elements.

PRACTICAL APPLICATIONS

Three radio telescopes have been built at Cambridge using the principle of aperture synthesis; all of them are transit instruments because of their size, and all are at frequencies low enough for the sensitivity to be adequate without repetition of observations with the particular configurations.

The first, at 38 mc, was constructed in 1954 by Blythe [1] who was able to synthesize a map of the sky as seen by a $3^\circ \times 3^\circ$ beam.

A similar instrument [5] having a pencil beam 47 minutes of arc between half-power points at 38 mc is now in operation. It is illustrated in Figs. 4 and 5. One element of the interferometer consists of a corner reflector 3300

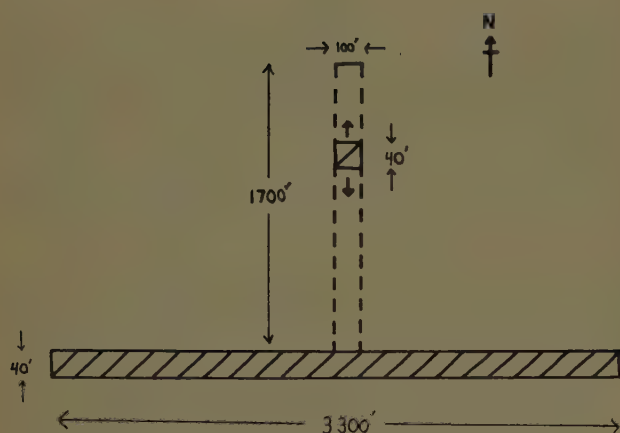


Fig. 4—Dimensions of the pencil-beam array at 38 mc.



Fig. 5—A view, looking east, of the fixed element of the 38 mc pencil-beam array.

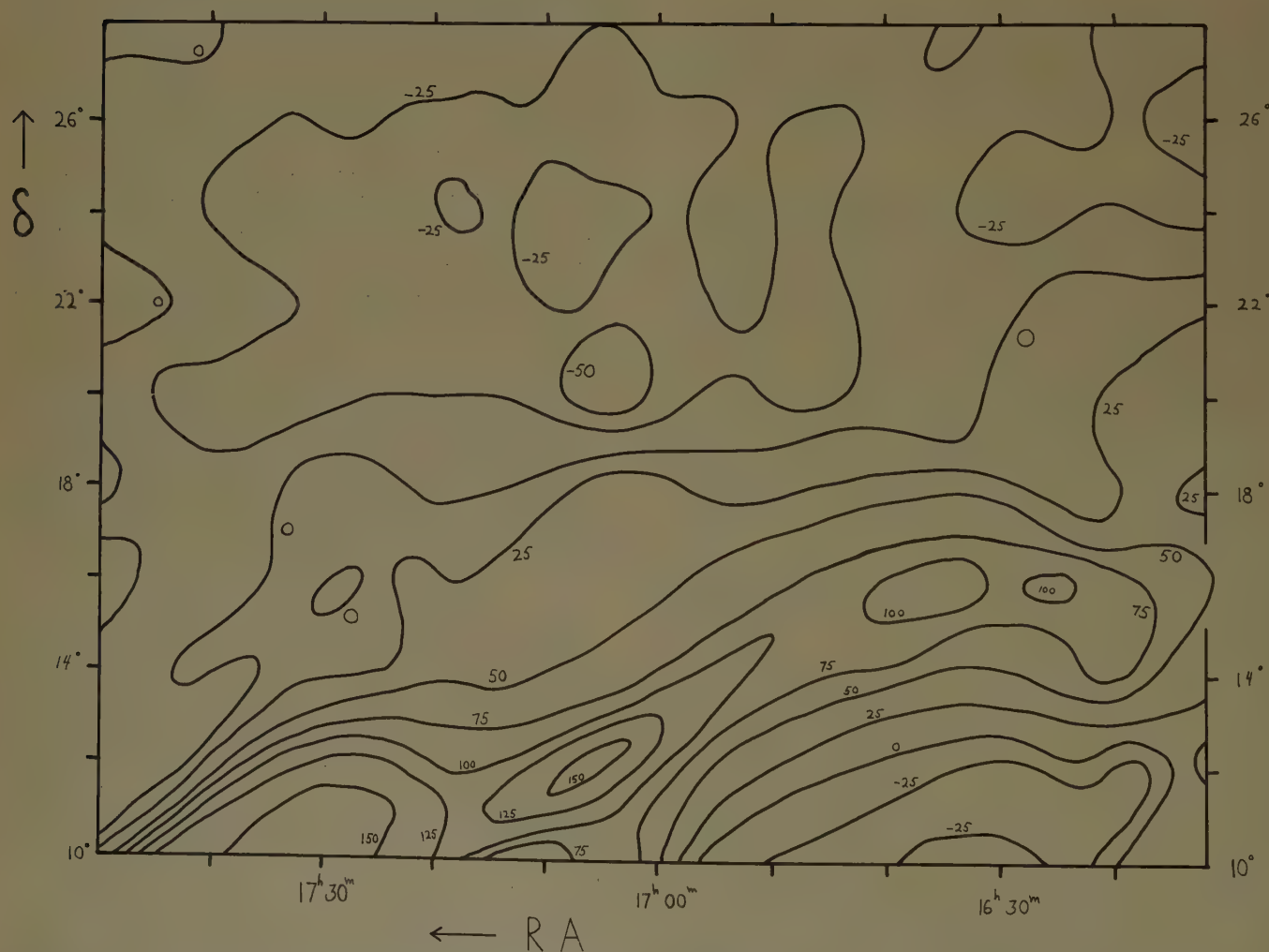


Fig. 6—A portion of a map of the sky as seen by the 38 mc array. It has been synthesized from observations made with spacings between the elements from 2λ to 13λ only.

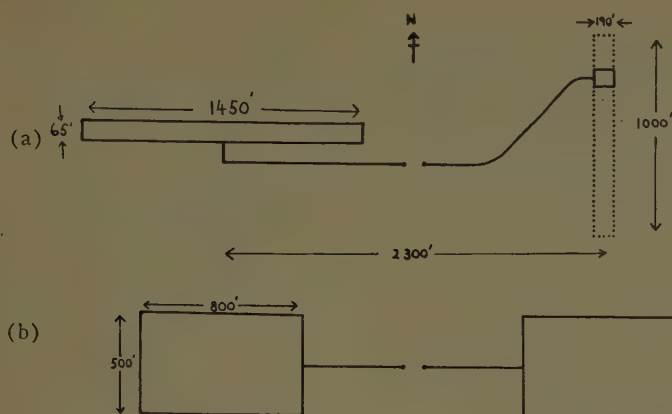


Fig. 7—(a) The dimensions of the 178 mc radio telescope.
(b) The equivalent interferometer.



Fig. 8—A view, looking west, of the fixed element of the 178 mc radio telescope.



Fig. 9—A view of the movable element of the 178 mc radio telescope.

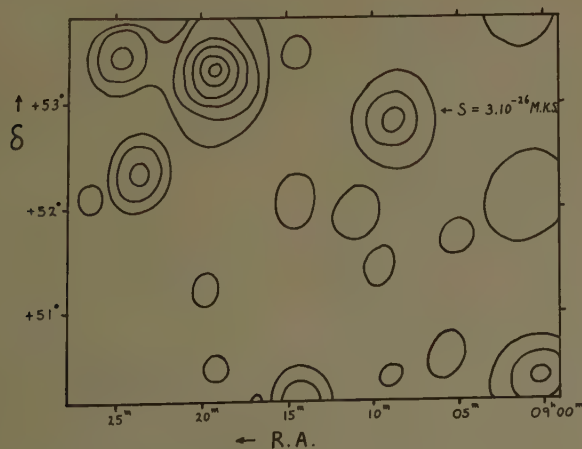


Fig. 10—A portion of a map of the sky at 178 mc showing some weak radio sources.

feet long and 40 feet wide. The movable element measures 100 feet by 40 feet and may be moved through a distance of 1700 feet along a north-south line. The instrument has a synthesized aperture equivalent in resolving power to one of sixty acres, and in sensitivity to one of one and a half acres. At the total cost of £3000, each square yard of equivalent aperture is worth about three cents. A portion of a map of the sky as seen by this telescope is shown in Fig. 6.

The other new instrument [5] is an interferometer for the study of radio stars at 178 mc. It is illustrated in Figs. 7, 8, and 9. The fixed antenna consists of an east-west array of dipoles 1450 feet long, supported at the line focus of a parabolic reflector 65 feet wide. The reflecting surface is constructed of horizontally strained wires supported by 34 towers and it may be directed, mechanically, to different elevations. The movable element, of similar construction, is 190 feet in length. Each of the six towers is mounted on rails and may be driven for a distance of 1000 feet along a north-south line.

The synthesized beamwidth is 18 minutes by 25 minutes of arc between half-power points, which is equivalent to the use of two apertures each measuring 800 feet by 500 feet. The cost was £50,000 or about 2 dollars per square yard of equivalent aperture.

Fig. 10 shows a map of a small portion of sky as seen by this instrument. The calculations have been carried out with the aid of EDSAC, the electronic computer at the Mathematical Laboratory, Cambridge.

To sum up, the advantages of the method of aperture synthesis are:

- a) economy—since less physical structure is required,
- b) the ease of altering the direction of the reception pattern in declination,
- c) flexibility—since the weighting factor can be readily changed in the calculations in order to adapt the beam for different purposes.

Features that may be disadvantageous are:

- a) the method requires the brightness distribution to remain unchanged during the period of observation and it is therefore of limited application in solar observations, and
- b) it is only efficient when surveying a comparatively large area of sky.

A detailed account of the radio telescopes and the methods of observation will be published elsewhere.

BIBLIOGRAPHY

- [1] J. G. Blythe, "A new type of pencil beam aerial for radio astronomy," *Monthly Notices Roy. Astron. Soc.*, vol. 117, pp. 644-651; 1957.
- [2] R. N. Bracewell and J. A. Roberts, "Aerial smoothing in radio astronomy," *Aust. J. Phys.*, vol. 7, pp. 615-640; 1954.
- [3] B. Y. Mills, A. G. Little, K. V. Sheridan, and O. B. Slee, "A high resolution radio telescope for use at 3.5 m," *Proc. IRE*, vol. 46, pp. 67-84; January, 1958.
- [4] M. Ryle, "A new radio interferometer and its application to the observation of weak radio stars," *Proc. Roy. Soc. (London) A*, vol. 211, pp. 351-375; 1952.
- [5] M. Ryle, "The Mullard Radio Astronomy Observatory, Cambridge," *Nature*, vol. 180, pp. 110-112; July, 1957.

Experimental Test of a Stepped Zone Mirror for Microwaves*

G. TORALDO DI FRANCIAT†, L. RONCHI†, AND V. RUSSO†

Summary—This paper is concerned with the construction and the experimental test of a stepped zone mirror, whose theory has been previously developed. According to the theory, this mirror is free from spherical aberration for any value of the angular aperture and at the same time is corrected for coma. The only monochromatic aberration which cannot be eliminated is astigmatism.

A model of the mirror has been built to operate with $\lambda = 3.2$ cm. The focal length is 86.4 cm and the angular aperture 120° . The test has given results in perfect agreement with the theory within a total field of about 40° . Special astigmatic receivers have been designed and built to eliminate the effect of the astigmatism. It is believed that the mirror may be used as a fixed antenna for radio astronomy.

INTRODUCTION

THE THEORY of a stepped zone mirror of the type considered here has already been developed.^{1,2} It is a microwave device, having the general shape of a spherical mirror with zonal steps. This mirror is essentially a diffraction grating.³ The introduction of the steps allows a perfect correction for spherical aberration and for the offense against the sine condition when the object (or the image) is at infinity. In agreement with optical theory, there is no coma⁴ of the first order with respect to the field angle.

A meridional section of such a mirror is shown in Fig. 1. The reflecting zones are sections of coaxial paraboloids, confocal at C . The intersections of the paraboloids with an ideal spherical surface Σ centered at C are circles whose centers, M_1, M_2, \dots lie on the symmetry axis CV . It is required that

$$VM_1 = M_1M_2 = M_2M_3 = \dots = n\lambda, \quad (1)$$

where n is an integer and λ the wavelength. This device is, of course, a "grating" of the generalized type.⁵ It may be shown² on the basis of "parageometrical" optics⁶ that when a plane wavefront normal to the axis impinges on the mirror, the n th order diffracted wave [where n is the

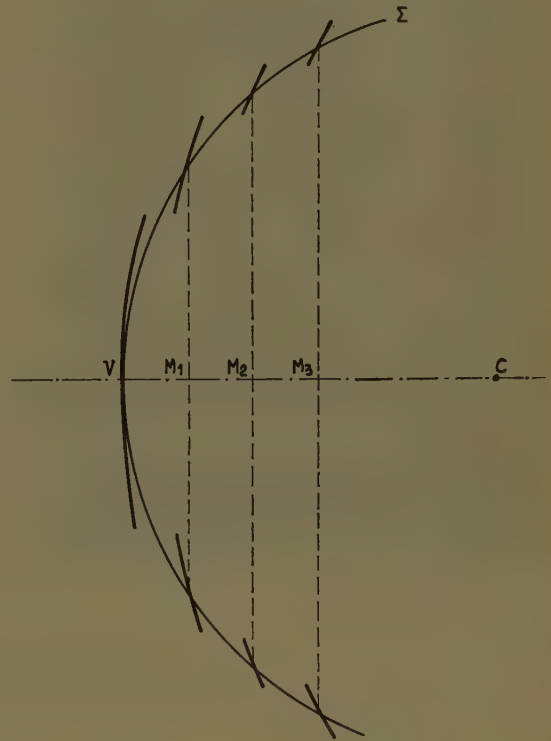


Fig. 1—Meridional cross section of the stepped zone mirror.

same integer appearing in (1)] is a perfect spherical wave centered at C . Moreover, it is a consequence of the theory of sawtooth gratings that the other diffracted waves are of negligible intensity.

The aberrations of the n th order wave were theoretically evaluated in terms of the field angle Ω and of the semiaperture θ .

A mirror of the type described has been built in order to test within what limits the theory is in agreement with experiment.

The mirror has the following specifications: focal length, $R = 86.4$ cm; angular semiaperture (as seen from C), $\theta = 60^\circ$; and intended wavelength, $\lambda = 3.2$ cm. Consequently, diameter $D = 150$ cm $\simeq 47\lambda$.

The points M_i appearing in Fig. 1 and (1) have been determined by putting $\lambda = 3.2$ cm and $n = 1$ (with this value of n , the chromatic aberration has the smallest possible value). The total number of zones is $N = 14$. The central zone and the first two zones around it are sections of paraboloidal surfaces, while the other zones have been realized as sections of conical surfaces; this approximation is justified because these zones are very narrow, being of the order of one wavelength.

The zones are made of an aluminum sheet, 1 mm thick and are supported by a wood frame (see Fig. 2). Consequently, the zones are insulated from one another.

* The research reported in this document has been sponsored in part by the Air Research and Development Command, U. S. Air Force, under Contract AF 61 (052)-67, through the European Office, ARDC.

† Centro Microonde, Florence, Italy.

¹ D. W. Fry and F. K. Goward, "Aerials for Centimetre Wavelengths," Cambridge University Press, Cambridge, Eng., p. 64; 1950.

² L. Ronchi and G. Toraldo di Francia, "An application of parageometrical optics to the design of a microwave mirror," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-6, pp. 129-133; January, 1958.

³ G. de Coligny and A. Fournier, "Physique et Technique des Systems Focalisants a Réseau Stationnaire," presented at the Congrès International Circuits et Antennes Hyperfréquences, Paris, France, pp. 8-11; 1957.

⁴ A mirror with four steps corrected for coma is described by J. F. Ramsey and J. A. C. Jackson in "Wide-angle scanning performance of mirror aerials," *Marconi Rev.*, vol. 19, pp. 119-140; July-October, 1956.

⁵ G. Toraldo di Francia, "Electromagnetic Waves," Interscience Publishers, Inc., New York, N. Y., p. 232; 1955.

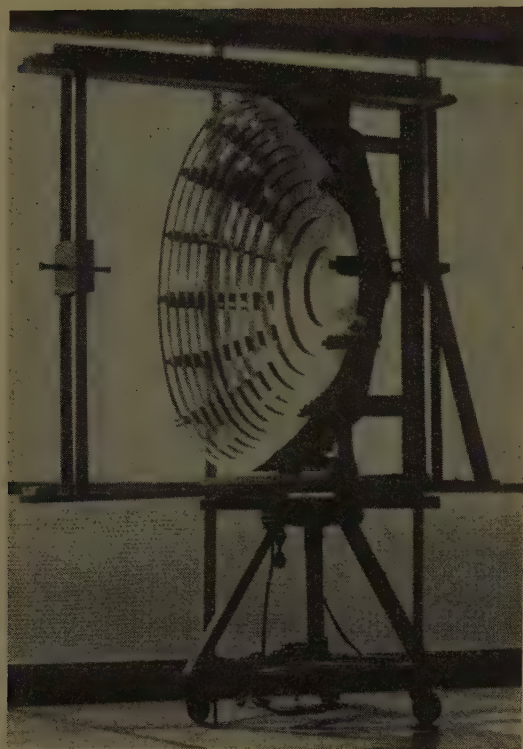


Fig. 2—The mirror on its platform.



Fig. 3—The feed.

The feed used is represented in Fig. 3. It is simply constituted by a rectangular waveguide, of standard dimensions, having at one end two small chokes. A light frame of wood and bakelite supports the feed in the desired (adjustable) position in front of the mirror.

WAVEFRONT TESTS

As a first step, the mirror was tested with a wavefront plotter described elsewhere.⁶

For a given value of the field angle Ω , the feed was placed, as shown in Fig. 4, in the horizontal plane containing the axis CV . The distance from the mouth F of the feed to V will be denoted by ρ and is of the order of R . The equiphase lines of the reflected field were determined in the same horizontal plane. By varying the value of ρ , the shape of these equiphase lines is varied. The value ρ_T of ρ , for which the equiphase line is closest to being a straight line, specifies the position of the tangential focus for the field angle Ω .

Fig. 5 shows the locus of the tangential foci determined in this manner, and also shows some reflected

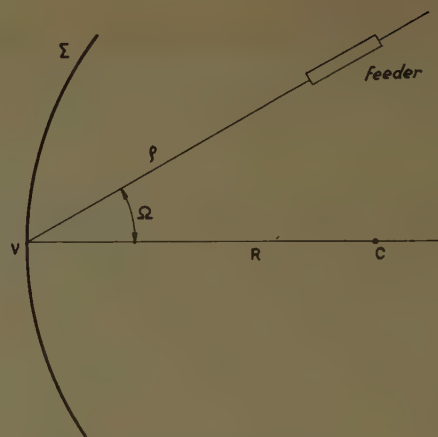


Fig. 4—The mirror and the feed.

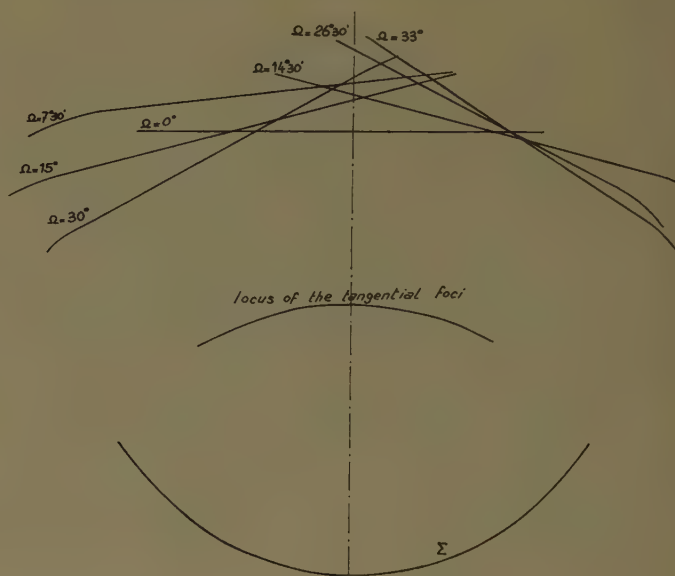


Fig. 5—Equiphase lines (wavefronts) emerging from the mirror.

wavefronts, varying from 0° to 35° for Ω , and for an electric field polarized in the horizontal plane. The wavefronts can be considered as straight lines for a great part of their extent. The "bends" appearing at the ends are to be ascribed to the diffraction caused by the finite dimensions of the mirror.

In order to perform a comparison with the theory, we recall (12), (13) of Ronchi and Toraldo di Francia.²

$$\rho_T = \frac{R \cos^2 \Omega}{2 \cos \Omega - 1} \quad (2)$$

This equation represents in polar coordinates a line which is shown in Fig. 6, where it is denoted by T . The comparison between theoretical and experimental values of ρ_T is made in Fig. 7, where, for convenience, ρ_T and Ω are used as rectangular coordinates. The experimental results are represented by crosses, and the theoretical predictions by a continuous line. The agreement is fairly good, within the limits of experimental accuracy, up to the value $\Omega \approx 20^\circ$.

⁶ P. F. Checcacci and V. Russo, "Prova sperimentale del funzionamento di alcune lenti di configurazione per microonde," Tech. Note 10, Contract AF 61 (514)-903, April, 1957, found in *Alta Frequenza*, vol. 27, pp. 92-107; April, 1958.

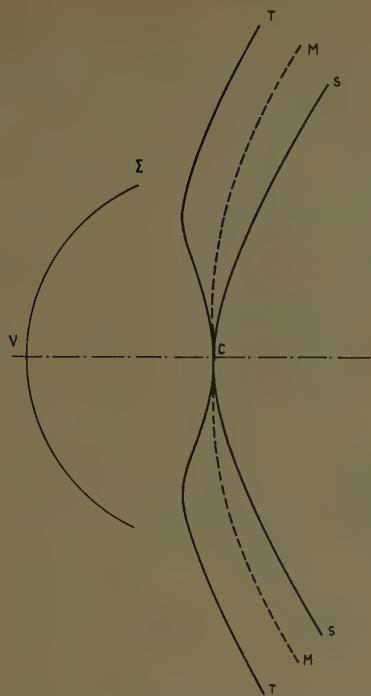


Fig. 6—Theoretical loci of the sagittal focus (S), tangential focus (T), and middle focus (M).

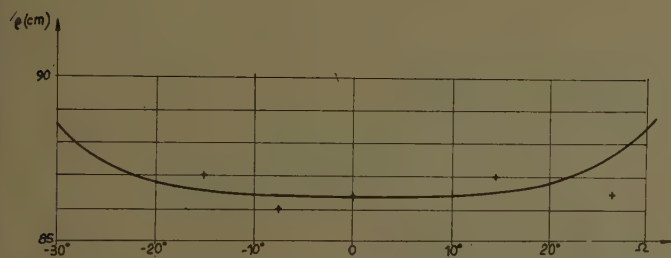


Fig. 7—Comparison between the theoretical locus of the tangential focus and that determined experimentally.

The locus of the sagittal foci has not been determined because of experimental difficulties.

RADIATION PATTERN TESTS

Another set of measurements has been carried out on the radiation patterns of the mirror for different field angles.

The setup for these experiments is that commonly used for plotting the radiation pattern of an antenna (see Fig. 8). It consists of a transmitter T which provides a plane wave impinging on the mirror Σ . The mirror is placed on a rotating platform P in such a way that the axis of the mirror is horizontal and remains horizontal during rotation. The incident "rays" are also horizontal. Following optical nomenclature, the horizontal plane containing the axis of the mirror will be termed the "meridional" plane. The rays belonging to the meridional plane meet at the tangential focus, while those belonging to a plane perpendicular to the meridional plane and through the "principal" ray meet at the sagittal focus.

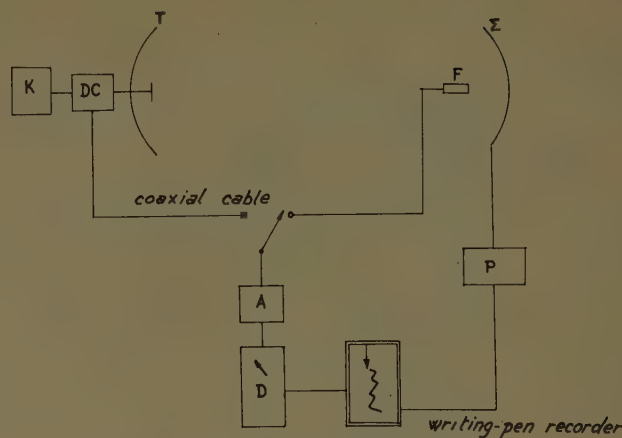


Fig. 8—Setup for plotting radiation patterns.

The mirror is equipped with a receiver F (the same waveguide used as feed in the preceding experiments) which can be moved both in the vertical and horizontal directions. The signal from the receiver is amplified and then detected.

The measurements are taken in the following manner. First the receiver is placed in the neighborhood of the meridional plane. Let Ω be the angle between the axis of the mirror and the secondary axis on which the receiver is located; let ρ be the distance from the receiver to the vertex of the mirror. In order to measure Ω , the base of the mirror is rotated until maximum signal is obtained; in this condition the direction of the incident wave makes the same angle Ω with the axis. The axial position ($\Omega = 0$) is determined by symmetry, *i.e.*, by requiring that the two positions $+\Omega$ and $-\Omega$ give the same result.

The receiver is then moved along a vertical line; the output will pass through a maximum when the receiver intersects the meridional plane. This maximum value obviously depends on the values of Ω and ρ . By varying ρ , one determines the distribution of the intensity along the secondary axis corresponding to the field angle Ω . In order to make this measurement more reliable, from time to time we have compared the signal from the receiver with a reference signal directly derived from the transmitter by means of a directional coupler (DC) and a coaxial cable. This reference signal is amplified and detected by the same instruments as is the signal from the mirror.

The results are shown in Figs. 9 and 10, which correspond to vertical and horizontal polarizations of the electrical field, respectively. Output power is plotted against ρ .

From these curves, we can derive the value $\bar{\rho}$ of ρ which, for a given value of Ω , corresponds to maximum intensity. The value of $\bar{\rho}$, with the experimental uncertainty, is plotted vs Ω in Figs. 11 and 12, for the vertical and horizontal polarizations, respectively. In the same figures, the theoretical curves of the middle focus ρ_M and of the tangential focus ρ_T are shown. The

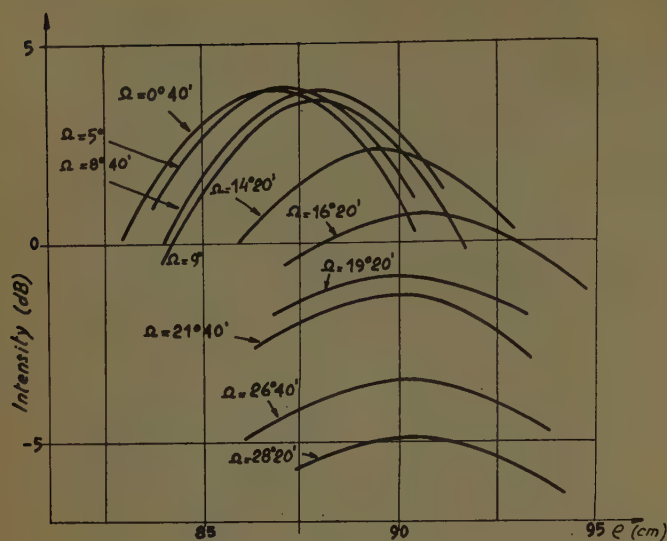


Fig. 9—Intensity plotted vs ρ for a number of values of Ω ; vertical polarization.

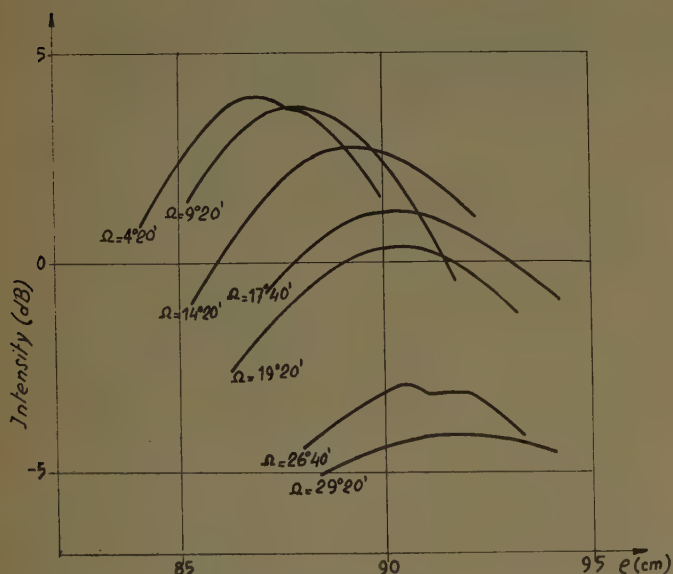


Fig. 10—Intensity plotted vs ρ for a number of values of Ω ; horizontal polarization.

values of $\bar{\rho}$ follow the curve of the middle focus well up to $\Omega \approx \pm 17^\circ$.⁷ For values of Ω larger than $\sim 17^\circ$, the experimental curve of maximum intensity bends toward the theoretical curve of the tangential focus.

Figs. 11 and 12 do not show any substantial difference between horizontal and vertical polarization of the electric field.

To determine the loci of the sagittal and tangential foci, it is useful to recall the geometrical optics description of astigmatism. All the rays of the beam first meet the tangential Sturm line, and then the sagittal Sturm line. The tangential Sturm line is (in our case) vertical and its middle point is the tangential focus, while the sagittal Sturm line is horizontal and its middle point is

⁷ This value of Ω is in a very good agreement with the results described in the preceding section.

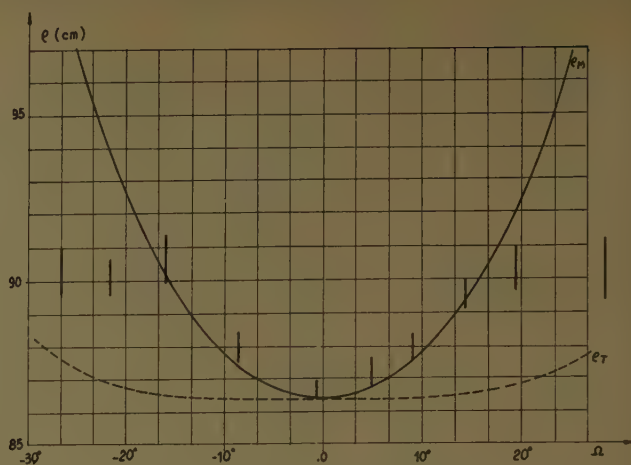


Fig. 11—Comparison between the experimental curve of best focus and the theoretical loci of tangential and middle focus (vertical polarization).

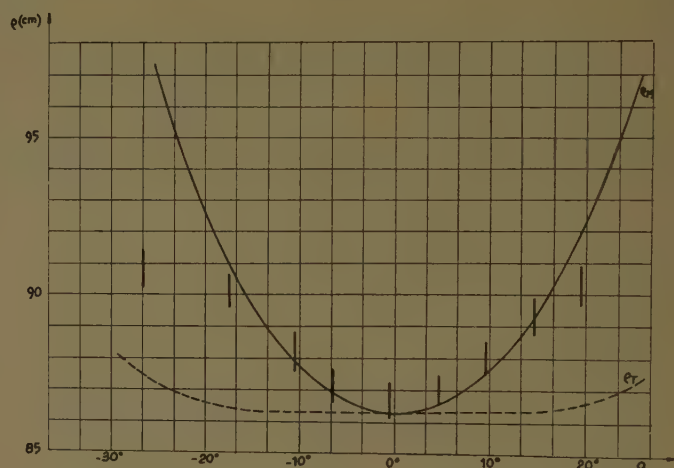


Fig. 12—Comparison between the experimental curve of the best focus and the theoretical loci of tangential and middle focus (horizontal polarization).

the sagittal focus. It is now clear that the tangential Sturm line will be located where the reflected beam has minimum width in the horizontal direction, while the sagittal line will be located where the beam has minimum width in the vertical direction.

The cross sections of the radiation pattern in the vertical plane have been determined, point by point, by moving the receiver along a vertical line for various values of ρ and Ω , and by plotting the intensity as a function of the distance from the point of maximum intensity. The measurements have been made for both vertical and horizontal polarization of the electric field. The results are shown in Figs. 13 and 14, respectively, where the half-power width is plotted against ρ for different values of Ω . Then, by plotting the value ρ_{\min} corresponding to minimum half-power width vs Ω , we obtain the experimental curve of the sagittal focus, as shown in Fig. 15, where the theoretical curve of the sagittal focus is plotted for comparison. There is good agreement between theory and experiment within the whole considered range of Ω .

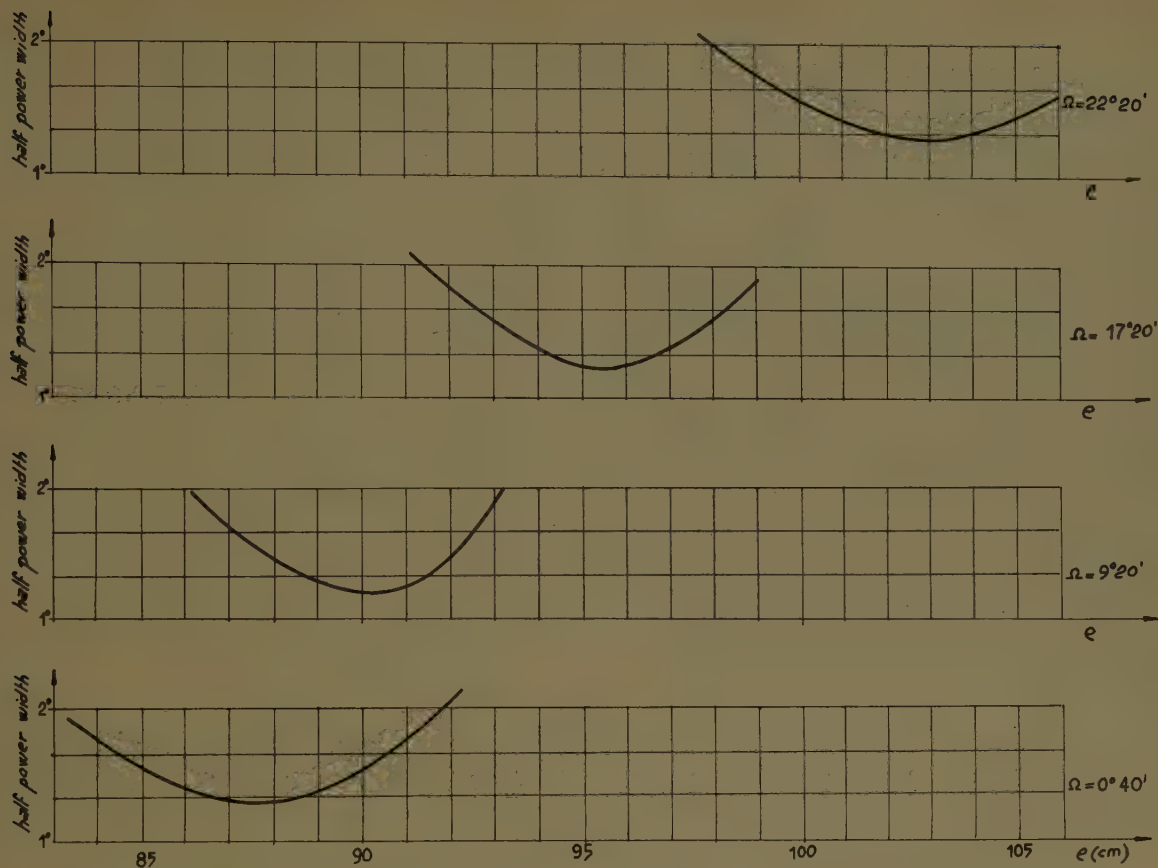


Fig. 13—Half-power width of the principal lobe in the vertical plane plotted vs ρ for a number of values of Ω (vertical polarization).

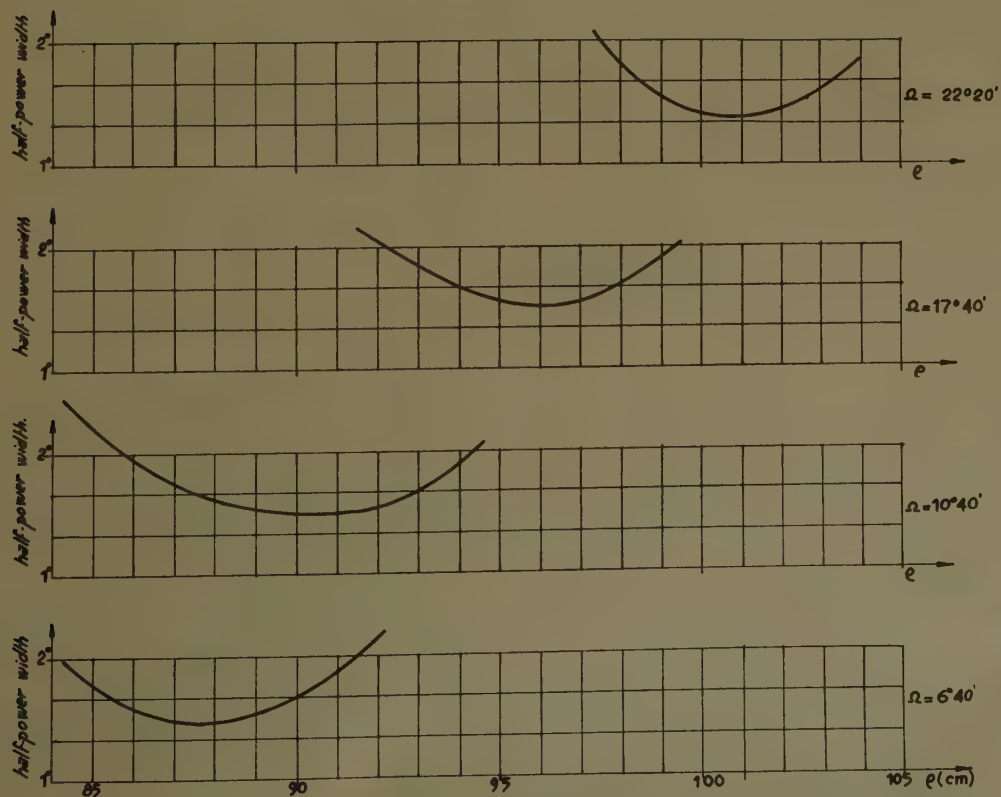


Fig. 14—Half-power width of the principal lobe in the vertical plane plotted vs ρ for a number of values of Ω (horizontal polarization).

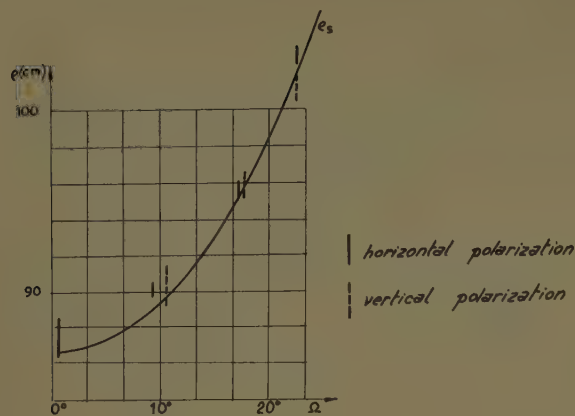


Fig. 15—Comparison between theoretical and experimental loci of the sagittal focus.

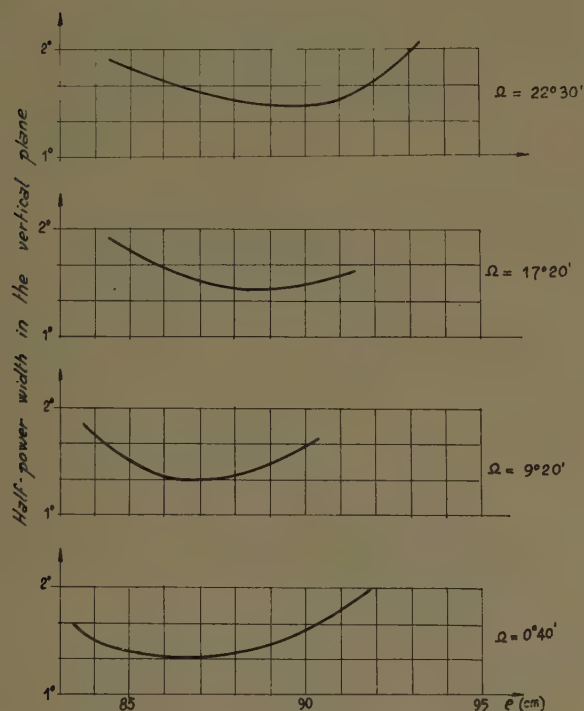


Fig. 16—Half-power width of the principal lobe in the horizontal plane plotted vs ρ for a number of values of Ω (vertical polarization).

The cross sections of the radiation pattern in the horizontal plane were determined by means of a pen recorder (see Fig. 8) coupled to the rotation of the platform. The half-power width was determined for a number of values of ρ and Ω , for both vertical and horizontal polarization of the electric field. The results are shown in Figs. 16 and 17, respectively. Then the values of ρ_{\min} corresponding to minimum half-power width (together with their uncertainty) were plotted vs Ω in order to obtain the experimental curve of the tangential focus. The comparison with the analogous theoretical curve is made in Fig. 18. It appears that the agreement is good up to the value $|\Omega| \approx 15^\circ$, after which the experimental line bends towards the theoretical line of the middle focus.

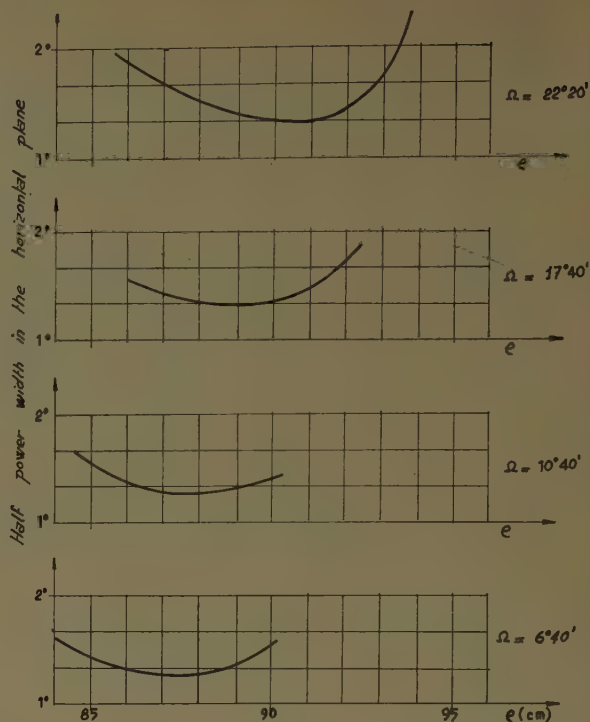


Fig. 17—Half-power width of the principal lobe in the horizontal plane plotted vs ρ for a number of values of Ω (horizontal polarization).

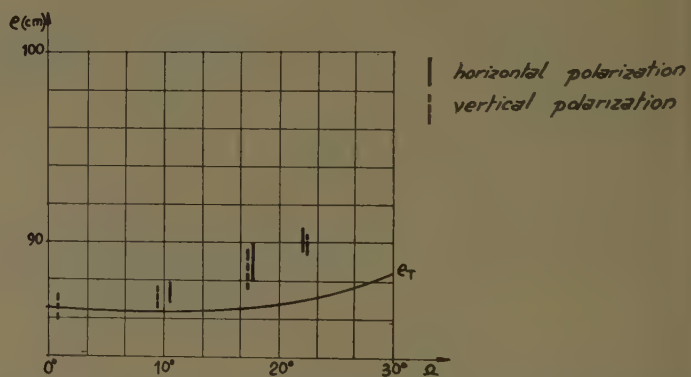


Fig. 18—Comparison between theoretical and experimental loci of the tangential focus.

CONCLUSION

In order to see the meaning of the obtained results, it will be useful to recall some results of the theory of diffraction of an astigmatic wave.⁸

So long as the wave aberration is very small (first approximation theory), the "best focus," or the point of the wave axis where the intensity has the highest value, coincides with the middle focus, *i.e.*, with the point midway between the tangential and sagittal foci. This coincidence is borne out very well by Figs. 11 and 12.

The question may arise concerning the stopping of this coincidence at an angle somewhere between $\Omega = 15^\circ$ and $\Omega = 20^\circ$. Now the total oscillation ΔW of the wave aberration for a given aperture and field can be evalu-

⁸ G. Toraldo di Francia, "La Diffrazione della Luce," Edizioni Scientifiche Einaudi, Boringhieri, Torino, Italy, pp. 476-482; 1958.

ated by means of (18) of Ronchi and Toraldo di Francia.² In our case, we have computed Table I, giving the values of ΔW in wavelengths for different

TABLE I

| Ω | 10° | 15° | 20° | 25° |
|--------------------|------|------|------|------|
| $\Delta W/\lambda$ | 0.39 | 1.00 | 2.02 | 3.56 |

values of Ω . The value of ΔW passes from 1 to 2 wavelengths precisely in the range 15°–20°. Therefore one should expect the first approximation theory to cease to be valid somewhere in that range.

Moreover, the dependence of the intensity on ρ , as shown in Figs. 9 and 10, is in agreement with the results of the first approximation theory (see Toraldo di Francia⁸) for small angles.

It would seem, therefore, that the behavior of the mirror, at least for small field angles ($\Omega < 20^\circ$), is in per-

fect agreement with that predicted by the parageometrical theory.

One possible discrepancy is that shown in Fig. 18. The position of the minimum half-power width in the horizontal plane does not coincide with the position of the tangential focus. However, we wish to emphasize the fact that, according to Ronchi and Toraldo di Francia,² there is some higher-order coma which affects the tangential, but not the sagittal Sturm line. In other words, all the rays pass *exactly* through the sagittal Sturm line, but only *approximately* through the tangential one. It therefore appears justified that the minimum half-power width in the vertical plane coincides with the sagittal line, while some discrepancy may be observed between the position of the minimum half-power width in the horizontal plane and the tangential line.

Finally, we want to emphasize the interesting result that very little, if any, difference is found between the results with vertical and horizontal polarization.

Preface to the Surface Wave Papers

JAMES R. WAIT†

IN VIEW of the current theoretical and experimental interest in electromagnetic surface waves, several sessions were devoted to this subject. In the present context, surface waves are regarded as any wave which glides along an interface between dissimilar media. With this in mind, active workers in the field were invited to submit papers representative of current research in this general area. In addition to the authors whose papers are contained herein, V. A. Fock, Val. Talanov, L. A. Weinstein, and M. A. Miller of the USSR and Z. Godzinski of Poland were invited to participate; unfortunately, they were unable to attend. Three other well-known authorities on surface wave phenomena, B. van der Pol of Holland, and N. Marcuvitz and F. J. Zucker of the U.S.A. were also not present; the former two because of illness and the latter because of the pressure of other work. The absence of our distinguished colleagues was keenly felt.

Following the sessions, a round-table discussion was held. The participants were chosen from the surface wave authors. The topics considered were: 1) influence of curvature; 2) implications of impedance boundary conditions; and 3) spatial modulation of the surface properties. No attempt was made at this time to set up a classification of surface wave types. The discussion, however, provided an excellent opportunity to air various viewpoints in the field.

For the next General Assembly of URSI, to be held in London in 1960, a summary report on the properties and classification of surface waves is to be prepared. Much of the material presented at the symposium is to be distilled and issued as a working document for distribution at the meeting. To assist me in this task are N. Marcuvitz (U.S.A.), F. J. Zucker (U.S.A.), L. A. Weinstein (U.S.S.R.), H. M. Barlow (U.K.), and A. L. Cullen (U.K.), who make up the other members of the working group appointed at Boulder at the last General Assembly.

† National Bureau of Standards, Boulder, Colo.

Anatomy of "Surface Waves"

S. A. SCHELKUNOFF†

Summary—The triple quotation marks in the title and elsewhere in this paper indicate that the enclosed words are used to denote the wave types enumerated in Table I. The purpose of the paper is to call attention to the fact that as a group these wave types have no important physical properties in common. Calling these wave types by the same name, even with qualifying adjectives, encourages one to assume that the most significant physical properties of one wave type are shared by other wave types. This has caused, and will continue to cause, we believe, serious misunderstandings. It is strongly urged that this loose use of words be abandoned. Another solution is possible. We could wait until everyone realizes that the term "surface waves" has become devoid of significant meaning and that it is equivalent to just "that thing."

INTRODUCTION

FOR obvious reasons the same word conveys different meanings to different individuals. Hence, some "noise" in communication between us is unavoidable. As long as the noise level is relatively low, we manage to understand each other reasonably well. When the noise level becomes high, serious misunderstandings are inevitable, and needless as well as wasteful controversies may arise. Such a situation has arisen in microwave theory in connection with the so-called "surface waves."

HISTORICAL BACKGROUND

It was Lord Rayleigh who discovered that in a semi-infinite elastic medium a source of finite dimensions excites two kinds of waves: 1) "space waves" which spread in all directions and 2) "surface waves" which spread along the boundary. If the medium is nondissipative, it follows from the principle of conservation of energy that at large distances from the source, the energy density in a space wave varies inversely as the square of the distance from the source and in a surface wave inversely as the distance. Surface waves seemed to be "attached" to the boundary of the solid and tended to follow it if it were curved.

In the time of Marconi's famous experiments and prior to the discovery of the Kennelly-Heaviside reflecting layer, there was much speculation about possible existence of similar kinds of electromagnetic waves. It was already known that electric waves had a tendency to cling to parallel wires ("Lecher wires," as they were called) and thus could be guided around corners. Did the surface of the earth have a similar tendency to capture some of the energy from an antenna and guide it into the shadow, thus explaining Marconi's success? That was the question. (See the Appendix.)

† Bell Telephone Labs., Murray Hill, N. J.

TABLE I

1. *Zenneck Surface Wave* (interface at half-space)
2. *Sommerfeld Surface Wave* (dipole overconducting half-space)
3. *Norton Surface Wave* (dipole over conducting half-space)
4. *Sommerfeld Axial Surface Wave* (imperfectly conducting cylindrical wire)
5. *Harms-Goubau Axial Surface Wave* (dielectric-coated wire)
6. *Plane Trapped Surface Wave* (dielectric-coated plane conductor, corrugated surface, or other inductive boundaries)
7. *Cylindrical Trapped Surface Wave* (same as above in cylindrical form)
8. *Plane Quasi-Trapped Surface Wave* (stratified conductor when the surface impedance has both a resistive and inductive component)
9. *Cylindrical Quasi-Trapped Surface Wave* (same as above in cylindrical form)
10. *Azimuthal Surface Waves* (on dielectric-coated and corrugated cylinders and spheres for propagation in the azimuthal direction)
11. *Composite Axial-Azimuthal Surface Waves* (same as above when propagation has a component in both the axial and azimuthal directions).

"SURFACE WAVES" SPROUT LIKE MUSHROOMS

In the half-century that followed the question, the answer had been at first "Yes," then "No," and a controversy began. At the same time many other types of "surface waves" appeared on the scientific scene, and at present one cannot be sure whether two different writers ascribe to their "surface waves" the same essential physical characteristics. It is a regrettable fact that clear communication among workers in the field does not exist. At a business meeting of Commission VI during the URSI General Assembly in Boulder, Colorado, a working group was formed to look into this matter and see what could be done. Dr. James R. Wait, chairman of the group, prepared the above preliminary list of wave types which at some time or another have been described by some writers as "surface waves."

INCIDENCE OF PLANE WAVES ON A PLANE BOUNDARY

It will be easier to understand the various kinds of waves in Table I in the light of propagation of plane waves in two semi-infinite, nonmagnetic, nondissipative media, separated by a plane boundary. Subsequently we shall consider the effect of dissipation and of curvature. To be more specific, let us assume that above the plane boundary is vacuum and below it a dielectric with the index of refraction

$$n_p = (\epsilon/\epsilon_0)^{1/2} \quad (1)$$

when there is no dissipation, and more generally the

"complex index of refraction"

$$n = \left(\frac{\sigma + j\omega\epsilon}{j\omega\epsilon_0} \right)^{1/2} = \left(\frac{\epsilon}{\epsilon_0} - j \frac{\sigma}{\omega\epsilon_0} \right)^{1/2} \\ = \left(n_p^2 - j \frac{\sigma\lambda\eta_0}{2\pi} \right)^{1/2} \quad (2)$$

where $\eta_0 = (\mu_0/\epsilon_0)^{1/2}$ and σ , ϵ , ϵ_0 have the usual meanings. Normally, plane waves incident on the boundary either from above or from below are partially reflected from it and partially transmitted across it. However, under certain circumstances, either the transmission or the reflection is total. Thus, if the H vector is *parallel* to the boundary there exists an angle of incidence for each medium, the "Brewster angle," when the transmission is total [Fig. 1(a) and 1(b)]. This case is of interest in connection with wave types 1 and 2 in Table I; *i.e.*, in connection with Zenneck and Sommerfeld wave types.¹ There is also an angle such that for all angles of incidence ϑ which are greater than the critical angle ϑ_c , the waves incident from below will be totally reflected *regardless of the orientation of the H vector* as shown in Fig. 2(a).

TOTAL TRANSMISSION ACROSS A PLANE BOUNDARY

Fig. 1(a) represents schematically the incidence at the Brewster angle from above. This angle is given by

$$\sin \theta_v = n_p(n_p^2 + 1)^{-1/2}. \quad (3)$$

The corresponding angle of refraction may be obtained from

$$\sin \theta_d = (n_p^2 + 1)^{-1/2}. \quad (4)$$

It is not surprising that θ_d is the Brewster angle for waves incident *from below*, and that the corresponding angle of refraction is θ_v . From this reciprocity it follows at once that in the case of a dielectric layer of *any thickness*, l , imbedded in vacuum [see Fig. 1(b)], the waves incident at the Brewster angle go through the layer without "seeing it." Similarly they will pass through a layer of vacuum in a dielectric medium without seeing it. Furthermore, the Brewster angle is *independent of the frequency of the waves*.

It is also clear that if the bottom boundary of the layer is a perfect conductor, the waves will be reflected only at this boundary, as shown in Fig. 1(c). For all other angles of incidence the reflection will be *multiple* and will take place at both boundaries.

Suppose now that the dielectric has a small conductivity σ and that

$$\sigma\lambda\eta_0 \ll 2\pi n_p^2, \quad (5)$$

¹ H. M. Barlow, "Surface waves," *PROC. IRE*, vol. 46, pp. 1413-1417; July, 1958.

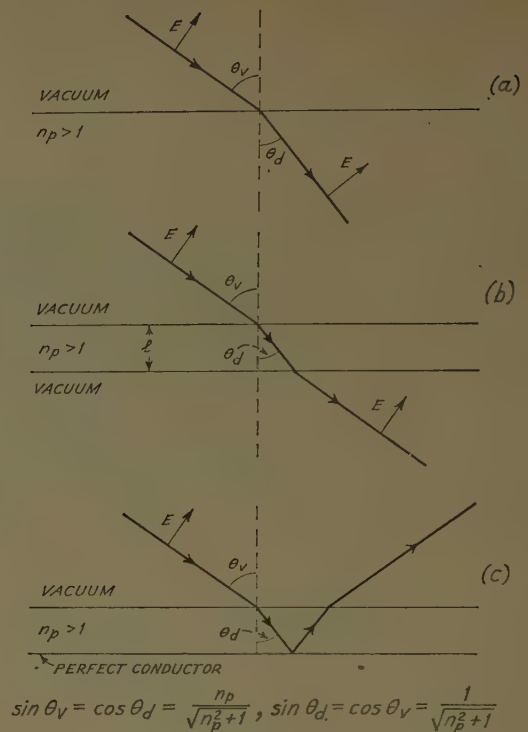


Fig. 1.

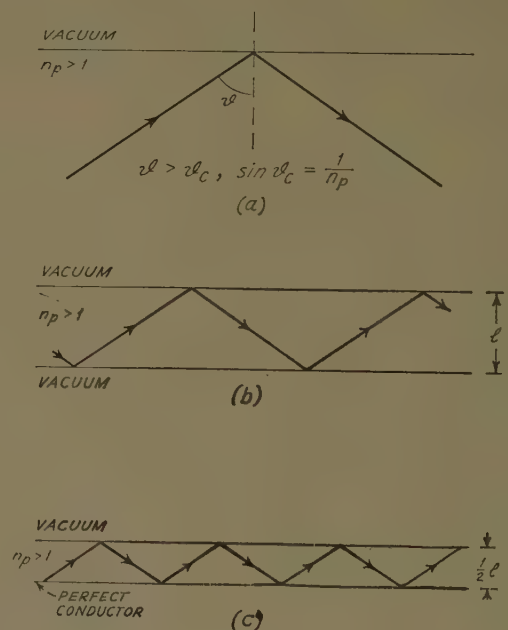


Fig. 2.

where λ is the wavelength in vacuum. This inequality is satisfied when the attenuation of plane waves passing through a layer of thickness equal to the wavelength in the dielectric is much smaller than π nepers. The effects of σ on the Brewster angle and the angle of refraction turn out to be of the second order. If the incident waves are *uniform plane waves*—that is, if their

amplitude is constant in any plane perpendicular to the rays—there will be small reflection. In this case the Brewster angle is the angle of minimum reflection. However, if the incident plane waves are inhomogeneous there exists a condition for no reflection. It is known that the rays of inhomogeneous plane waves in a vacuum are straight lines along which the phase change is maximum and the amplitude remains constant. The planes perpendicular to the rays are equiphase planes. In these planes there exists one direction in which the amplitude does not vary, while in the perpendicular direction the amplitude change is maximum. The condition of no reflection would occur when the attenuation constant in this direction is²

$$\alpha = \frac{1}{2}\sigma\eta_0(n_p^2 + 1)^{-3/2} \csc \theta_v. \quad (6)$$

The exponential rate of attenuation with the increasing distance from the interface (upward) is $\alpha \sin \theta_v$. Let us note that α is small and independent of λ [subject to condition (5)]. Let us also note that if condition (5) is satisfied for a certain wavelength, it is satisfied for all shorter wavelengths.

The inhomogeneous plane waves continue to be attenuated as they pass through the dielectric medium. In this case some of this attenuation is due to dissipation of energy in heat. In the case of a layer of finite thickness, the waves below the layer *will grow in amplitude* in the downward direction. This is evident at once from the fact that in vacuum the amplitude along the rays is constant. Likewise, in the case of reflection from a perfectly conducting boundary [Fig. 1(c)], the reflected waves above the layer will grow in the upward direction.

TOTAL REFLECTION AT A PLANE BOUNDARY

We shall now turn to the case of total internal reflection (Fig. 2). The critical angle is given by

$$\sin \vartheta_c = 1/n_p, \quad (7)$$

and is independent of the frequency. For all greater angles the reflection is total. The waves above the plane are evanescent and the attenuation constant in the vertical direction is

$$\alpha = (2\pi/\lambda)(n_p^2 \sin^2 \vartheta - 1)^{1/2}. \quad (8)$$

Note that α increases with the frequency, while in the case of Brewster incidence α , as given by (6), it is independent of the frequency.

In the case of a dielectric layer of finite thickness l [Fig. 2(b)], the total internal reflection will take place at both boundaries if it takes place at one. For most angles of incidence there will be a destructive interference. It is only for a certain infinite but discrete set of

angles that the interference is additive and not destructive. The waves become "trapped." For vertical polarization the "characteristic angles" may be obtained from Schelkunoff's equations,³ where the complete solution is given for either plane or radial cylindrical waves. For our purposes we need only his equations beginning with (21)-(23). Eqs. (21)-(27) determine all the characteristic or "trapped modes" and the corresponding angles of incidence. One set of trapped modes is not affected by an insertion of a perfectly conducting plane in the middle of the layer [Fig. 2(c)]. The angles of incidence for this set may be expressed in terms of a numerical parameter \hat{p} as follows:

$$\sin \vartheta = n_p^{-1} [1 + n_p^{-2} \tan^2 (\hat{p}/2)]^{1/2} \cdot [1 + n_p^{-4} \tan^2 (\hat{p}/2)]^{-1/2}. \quad (9)$$

The various modes correspond to the following ranges of \hat{p} : $0 \leq \hat{p} \leq \pi$, $2\pi \leq \hat{p} \leq 3\pi$, \dots , $2n\pi \leq \hat{p} \leq (2n+1)\pi$. The parameter \hat{p} is related to the ratio of the thickness of the dielectric layer to the wavelength

$$l/\lambda = (2\pi)^{-1}(n_p^2 - 1)^{-1/2} \hat{p} [1 + n_p^{-4} \tan^2 (\hat{p}/2)]^{1/2}. \quad (10)$$

The second set of trapped modes is determined by the same equations with "cot" appearing instead of "tan."

Let us consider in particular the first mode corresponding to $0 \leq \hat{p} \leq \pi$. As \hat{p} varies from 0 to π , λ varies from infinity to zero [see (10)]. From (9) we find that $\sin \vartheta$ varies correspondingly from $1/n_p$ to unity. Hence ϑ varies from the critical angle ϑ_c to $\pi/2$. The rate of evanescence of the waves in vacuum [see (8)] varies from zero to $(2\pi/\lambda)(n_p^2 - 1)^{1/2}$. That is, the waves in vacuum become more rapidly evanescent as λ decreases. As λ decreases, the energy associated with the waves is drawn more and more into the layer. An increase in l is equivalent to a decrease in λ .

If the dielectric layer is slightly dissipative, there will be a slight attenuation in the direction of propagation. This will be a first-order effect. The effect on the characteristic angle of incidence ϑ will be a second-order effect. Likewise, the effect of σ on the rate of evanescence of waves in vacuum will be a second-order effect (for well-trapped waves). The first-order effect on waves in air will be a small phase constant, so that some energy in the airwaves will pass into the layer. For well-trapped waves, however, the energy that is dissipated in the layer is already there.

CONTRAST BETWEEN TOTAL TRANSMISSION AND TOTAL REFLECTION

Thus the situations shown in Figs. 1 and 2 are radically different when the dielectric layer is nondissipative or slightly dissipative. In the first case the waves in

² The effect of σ on the Brewster angle of incidence is of the second order.

³ S. A. Schelkunoff, "Electromagnetic Waves," D. Van Nostrand Co., Princeton, N. J., pp. 429-430; 1943.

vacuum are plane, or for small σ they are essentially plane, and pass through the layer without seeing it. The Brewster angles in vacuum and in the dielectric are substantially independent of λ , subject to (5) which is true for sufficiently short waves. The phase constant in vacuum, $2\pi/\lambda$, increases as λ decreases, while the rate of evanescence given by (6) is small and independent of λ and the thickness of the layer. In the second case the waves in vacuum are always evanescent, and the rate of evanescence increases as λ decreases or as the thickness l of the layer increases. In the former case, one might say that a small σ introduces an "anomalous attenuation" at right angles to the direction of propagation of waves in vacuum. In the latter case, a small σ introduces an "anomalous phase constant" (and anomalous velocity of propagation toward the layer) into otherwise purely evanescent waves. Also, in the former case the waves merely pass through the layer, and in the latter they are trapped by it.

RADIAL WAVES INCIDENT ON A PLANE BOUNDARY

The case of radial waves is no different from the others. In free space an infinite current filament carrying progressive waves generates radial waves. Near the filament the field is largely reactive, but at greater distance the "rays" will make a constant angle with the filament. If a dielectric layer is introduced perpendicularly to the filament, there will generally be reflections at both boundaries of the layer as well as transmission through it. However, if the angle which the rays make with the axis equals the Brewster angle, there is no reflection and the transmission is total (Fig. 3). On the other hand, if this angle is larger than the critical angle of total internal reflection and if it is given by (9) for the wavelength determined by (10), then the waves will be trapped. In this case the current in the filament outside the layer should be attenuated with the increasing distance from the boundaries in accordance with (8), in order to make the wave a pure trapped wave (Fig. 4).

The effects of small dissipation in the dielectric are the same as for plane waves.

RADIAL WAVES INCIDENT ON A CYLINDRICAL BOUNDARY

The case of radial waves incident on a cylindrical boundary is essentially the same. In Fig. 5 we have an electric current filament on the axis of a dielectric rod. For a certain phase velocity of current we shall have total transmission. For certain other velocities there may be, depending on the ratio of the diameter to the wavelength and the index of refraction, total *in-phase* reflection (Fig. 6). In such cases the intensity of the field would become infinite for a finite current, and a finite field can be excited by a dipole.

The case of a dielectric-coated wire is essentially that of a portion of a dielectric-coated plane [Fig. 2(c)] folded into a cylinder.

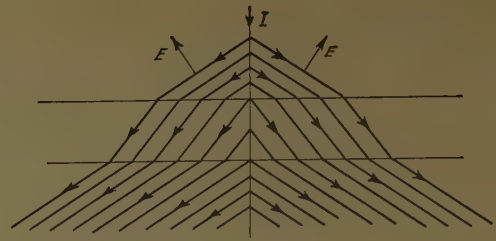


Fig. 3.

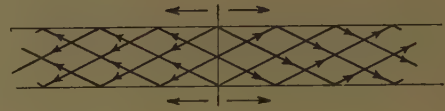


Fig. 4.

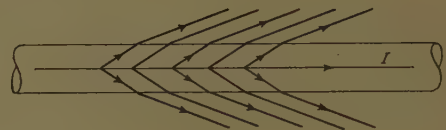


Fig. 5.

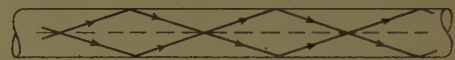


Fig. 6.

HIGHLY CONDUCTING LAYERS

For the present purposes we shall call a dielectric layer "highly conducting" if

$$\sigma\lambda\eta_0 \gg 2\pi n_p^2, \quad (11)$$

in contrast to (5). This condition is satisfied even for small σ if λ is sufficiently large. For metals it is satisfied for all radio frequencies.

When the inequality (11) is satisfied, the real part of the Brewster angle is

$$\theta_0 = \frac{\pi}{2} - (\pi/\sigma\lambda\eta_0)^{1/2}, \quad (12)$$

or nearly 90° . The attenuation of the inhomogeneous wave in vacuum is small:

$$\alpha = 2\pi(\pi/\sigma\lambda^3\eta_0)^{1/2}. \quad (13)$$

In this case, the real part of the Brewster angle is seen to depend on the conductivity and the wavelength. Otherwise the situation is not different from that existing in the case of nondissipative and slightly dissipative layers. Fig. 7 shows schematically a radial wave incident at the Brewster angle on a metal plate with a hole in it. Of course, the angle the incident waves make with the plate is greatly exaggerated. From the point of view of geometrical optics, the plate has no effect on the wave generated by the filament—there is no "back-fire" into the upper region except for the inevitable diffraction around the edge of the hole.

For an elevated dipole (Fig. 8), the cone of rays mak-

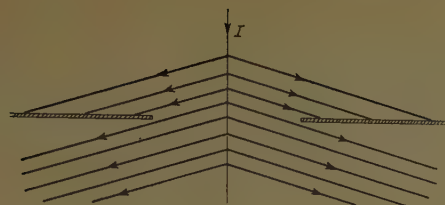


Fig. 7.



Fig. 8.



Fig. 9.

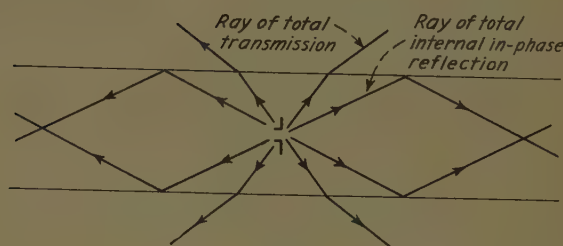


Fig. 10.

ing the angle θ_v with the axis of the dipole traces on the plane a "boundary circle" separating the region where the reflected waves reinforce the incident waves from the region where they weaken them. At the circle the reflection coefficient is minimum (0.41⁺). Incidentally, (12) gives the real part of the "Brewster angle" for the inhomogeneous waves. In the case of the dipole the incident waves in a limited region are substantially uniform, and the angle of minimum reflection, if we still call it θ_v , has (for metals) a factor of two under the square-root sign in (12). These details, however, are unimportant for our purposes.

If the dipole is right over the plate, as in Fig. 9, the pencil of rays making the angle θ_v with the axis again separates the region in which the image of the dipole reinforces the free space wave of the dipole from the region where it tends to destroy it. The effect of the image extends to the surface itself. Thus, near the dipole the field will vary inversely as the distance from the dipole, while "far away" it will vary inversely as the square of the distance. The "far away zone" should somehow be related to $\frac{1}{2}\pi - \theta_v$. An order of magnitude may be determined from the conditions prevailing in the case of an elevated dipole. The radius of the circle of minimum reflection depends on the height of the dipole and on $\frac{1}{2}\pi - \theta_v$. But it may be conjectured that when the height becomes less than $\lambda/2$, or possibly $\lambda/4$, the radius becomes substantially constant. Hence, "far away" means at distances much greater than

$$d = (\sigma\lambda^3\eta_0/8\pi)^{1/2}. \quad (14)$$

Little needs to be said about "trapped" waves in metallic media. They are attenuated extremely rapidly.

A DIPOLE IN A NONDISSIPATIVE OR SLIGHTLY DISSIPATIVE LAYER

Let us now return to a nondissipative dielectric layer, and assume a dipole inside it (Fig. 10). In an infinite space, the dipole pattern is a toroid. If the layer is very thick compared with the wavelength, we could determine the pattern, more or less, with the aid of geometrical optics.

Waves traveling in directions within the "Brewster cone" will be partially transmitted and partially reflected at each boundary. Along the Brewster rays the transmission will be total. In directions outside the Brewster cone but inside the critical cone of total internal reflection, the waves will again be partially transmitted and partially reflected. In the directions outside the critical cone, the reflection at each boundary will be total. In particular, in directions surrounding those given by (9) and (10), there will be a cumulative effect due to the right phase relationships, and normally spherical annual beams will be converted into radial beams, or trapped beams.

The problem can be formulated exactly with the aid of Sommerfeld integrals for the dipole field.⁴ From the exact formulation we learn that the representation of the spherical wave in terms of plane waves includes inhomogeneous plane waves as well as uniform plane waves traveling in all directions. The presence of the inhomogeneous waves which are attenuated in the vertical direction and whose velocities in the radial direction are less than the velocity of uniform plane waves, accounts for the fact that trapped waves can be "stirred up" by a dipole above the layer—a fact incomprehensible in the light of pure geometrical optics. The stirring up of trapped waves by a dipole is analogous to the stirring up of free oscillations of a system by an impressed impulse of force. Mathematically both phenomena are associated with the poles of the reflection coefficient (or certain impedance functions). The Brewster effect is associated with the zero of the reflection coefficient or the "matching" of characteristic impedances.

SEMITRAPPED WAVES ON A PERFECTLY CONDUCTING WIRE

In free space, electric and magnetic fields are generally diffuse. In the vicinity of a thin charged wire, on the other hand, the electric field is highly concentrated,

⁴ *Ibid.*, pp. 413-417, 428-431.

as is the magnetic field in the vicinity of a thin wire carrying current. Hence, we suspect that a thin wire will exhibit a tendency to trap electromagnetic energy and guide it. This is indeed the case. Fig. 11 shows schematically a quarter-wave antenna connected to one terminal of a generator and a semi-infinite wire to the other. Some of the energy will be radiated away and some trapped by the semi-infinite wire. However, if the wire is perfectly conducting and if its radius is finite, the current at large distances from the generator will vary inversely as the logarithm of the ratio of the distance to the radius of the wire. Thus, the energy in the vicinity of the wire will be slowly diminishing, and the field will be spreading farther and farther from the wire.

Even a thin layer of the dielectric coating will convert this semitrapped wave into a fully trapped wave. On the other hand, the resistance in the wire will have the opposite effect. In addition to the expected loss of energy in heat, at large distances from the generator there will be a tendency for the wave to break away from the wire as in the case of spherical waves over an imperfect ground.

In this case of an imperfectly conducting wire (uncoated), the current will vary at first more or less as it does in a perfectly conducting wire (there will be some attenuation); but at very large distances, the current and the field around the wire will vary inversely with the distance from the generator. Thus imperfectly conducting wires eventually lose their "grip" on the energy.

CONCLUSION

We have discussed at some length the behavior of plane and radial waves at plane and cylindrical boundaries across which the index of refraction is discontinuous. We have been concerned, in particular, with the conditions of total transmission and total reflection. At the Brewster angle of incidence the vertically polarized waves are totally transmitted. If the waves are not vertically polarized, the reflected waves will be horizontally polarized since the other component will be totally transmitted. Sharp spherically divergent beams aimed in the Brewster direction will be largely transmitted across the boundary, with some reflection. The same would be true of beams falling on a spherical boundary of large radius. Sharp beams aimed in the Brewster direction at a large, dielectric-coated, perfectly conducting sphere will generally pass through the dielectric and will be reflected from the sphere—but there will be small multiple reflections off the axis of the beam.

We have also examined the case of total internal reflection and the consequent trapping of waves. This occurs in the dielectric layers or cylinders when the waves are incident on the boundary from the dielectric side at certain angles larger than the critical angle of

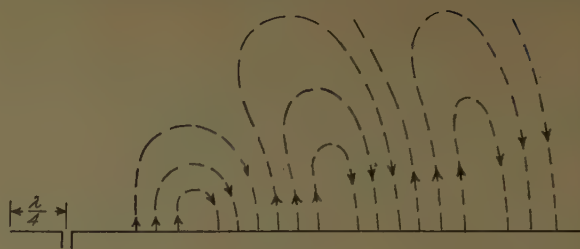


Fig. 11.

total internal reflection. When the index of refraction is large, this angle is only slightly larger than the Brewster angle. Hence, even fairly narrow spherical beams aimed at the Brewster angle may excite trapped waves, although a considerable fraction of power will be transmitted across the boundary.

Trapped waves follow slow bends in dielectric layers and rods, although some energy escapes.

Referring to Table I in this paper, we note that the wave types 1, 2 and 4 are associated with the Brewster angle of incidence. Wave types 5 through 11 are trapped waves. In highly conducting dielectric layers and cylinders, the trapped waves attenuate so fast that they represent merely local effects.

Wave type 3, the Norton "surface wave," is not related to any of the above types. This wave was defined as the difference between the exact field of a dipole above an imperfect ground, and the field calculated by the rules of geometrical optics. This "wave" does not satisfy Maxwell's equations.

There is also the term "ground wave" which is used by radio engineers to denote the *total* wave which would have existed on the ground surface if the Kennelly-Heaviside layer were absent. The wave reflected from the layer is called the "sky wave." In the primary service area for all broadcasting stations operating in the low and medium frequency ranges, the sky wave is very weak and only the ground wave is important. This ground wave has some relation to the Norton "surface wave," but is not identical to it. The latter vanishes for a perfect ground when the ground wave of the radio engineer is the strongest. The ground wave has also been confused with the Zenneck "surface wave."

It is this writer's opinion that the loose use of the term "surface wave" is unfortunate and causes a great deal of unnecessary confusion. If it is continued, the best that one could hope for is that the term will become entirely devoid of meaning. This writer hopes, however, that the classical definition of the term (Lord Rayleigh's) will be restored. Sommerfeld and Zenneck adhered to it (see the appendix below), although they have made an unfortunate slip in their analysis which subsequently confused the issue. In the classical sense, the term "surface wave" applies only to wave types 5 through 11 of all those mentioned in Table I.

APPENDIX

The following excerpt from Zenneck's "Wireless Telegraphy"⁵ shows that Sommerfeld and Zenneck accepted Rayleigh's definition of "surface waves" as far as the most significant physical properties are concerned. Zenneck writes:

Instead the facts, according to A. Sommerfeld's theory, are as follows:

a. Surface and Space Waves.—The waves which emanate from a transmitter placed in a homogeneous insulating material were discussed in Art. 20. They are characterized by the fact that energy is radiated in straight lines, radially from the transmitter. Consequently, the energy varies as $1/r^2$ (r =distance from source) and the amplitudes of the electric and magnetic field strengths vary as $1/r$. We will refer to these as "space waves" in what follows.

A different kind of wave is obtained, e.g., with Lecher's system [Art. 72c]. Here the waves travel along the wires, following any bends they may have. The flow of energy along the wires and the amplitude of the waves would remain constant during their progress, were it not for the fact that a portion of the energy is consumed in the wires (due to Joulean heat developed). This causes a gradual reduction in the energy and wave amplitude along the course of travel, a phenomenon which is termed "absorption." We will refer to waves of this kind as "surface waves," as they follow the surface of the conductor.

b. The wave emanated into the air by an antenna at the earth's surface may be conceived as consisting of two component parts, one of which is of the nature of a space wave, the other of a surface wave. In the former the energy $\propto 1/r^2$, the amplitude therefore $\propto 1/r$; in the latter the energy $\propto 1/r$, the amplitude $\propto 1/\sqrt{r}$. The fact that in the latter there is a decrease in the energy as the distance increases, in contrast to the wave following a wire—and in addition to and entirely aside from such absorption as occurs—is explained by the fact that the energy is spreading itself out over ever-increasing circles, as the wave travels its course.

Absorption of course occurs in addition to this reduction in amplitude due to the expansion of the wave in space. As each wave advances through the air it is accompanied by a wave in the ground. And as the ground always has more or less conductivity, the moving electric field, constituting the wave, results in the formation of currents, just as in the wires of the Lecher system. These currents consume energy, which is drawn from that of the waves radiated by the antenna, so that an absorption occurs in this way.

This much is relevant to the classical distinction between space and wave types. The next excerpt led to a controversy and contributed to subsequent confusion. He continues:

c. While at short distances from the transmitter, the waves are almost entirely of the nature of space waves, as the distance increases the surface component becomes more and more predominant, as its amplitude decreases more slowly than that of the surface component.

⁵ J. Zenneck, "Wireless Telegraphy," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 249–250; 1915.

That is, the nature of the wave constantly approaches that of a surface wave. When the distance becomes very great, the surface wave may again give way to the space wave, as the former is more rapidly absorbed. It is questionable, however, whether this effect is of practical importance.

This change is the more rapid, the shorter the wave-length is and the lower the conductivity and dielectric constant of the ground are. A calculation of the distance at which the actual amplitude of the wave differs by 10 per cent from the amplitude of the space wave, results in the following figures:

| | | |
|--------------|---------------------|-------------------------------|
| Sea water | $\lambda = 2$ km. | Distance = 20,000 km. approx. |
| Sea water | $\lambda = 1$ km. | Distance = 5000 km. approx. |
| Sea water | $\lambda = 0.3$ km. | Distance = 500 km. approx. |
| Fresh water. | $\lambda = 2$ km. | Distance = 4 km. approx. |

The distance becomes still shorter with dry ground.

Hence, while with sea water for all distances which come into consideration—20,000 km. is half the circumference of the earth—and for all wave-lengths over 1 km. the waves have the characteristics of space waves, with fresh water and even far more so with dry ground, they assume the characteristics of surface waves at distances of only a few wave-lengths or even less than one wave-length. Hence the nature of the wave propagation in this case must not be conceived as being the same as that described in Art. 138 over sea water.

The above quoted conclusion of Zenneck's is based on the original formulas obtained by Sommerfeld. Burrows⁶ has pointed out that numerically the transmission formulas based on Sommerfeld's results differ from those of Weyl by just the surface wave term P , and has made careful measurements which support the results of Weyl.

Subsequently, it became clear that the source of the difficulty involving the surface wave term in Sommerfeld's formulas was the double-valued nature of the square root terms in the reflection coefficient appearing in the integrand. In the mathematical formulation of the *physical* problem, it is essential that the square roots be assigned the values whose real parts are positive. Subsequent deformation of the contour of integration has to be conducted with great care and circumspection. No difficulty would have arisen if the deformation were made in a complex plane with an impassable cut so that the square roots could take on only their principal values. As it happened, the deformation was made on a Riemann surface where it is quite easy for the reflection coefficient in the integrand to turn into its reciprocal.

⁶ C. R. Burrows, "Existence of a surface wave in radio propagation," *Nature*, vol. 138, p. 284, August 15, 1936; also, "The surface wave in radio propagation over plane earth," *PROC. IRE*, vol. 25, pp. 219–229; February, 1937.

Waves on Interfaces

GEORG GOUBAU†

Summary—In wave propagation along plane or cylindrical interfaces a distinction is made between truly guided waves (surface waves) and partially guided waves (radiating waves). When dissipation losses are involved, surface waves no longer represent the asymptotic field near the interfaces at large distances from the source. In order to separate them from the total field it is necessary to have a criterion for radiating waves which are free of any surface wave components. Such a criterion exists in the form of orthogonality relations which are the mathematical formulation of the concept that an ideal antenna which excites only a surface wave should not receive a radiating wave.

I. INTRODUCTION

WAVE propagation along interfaces between different media was a controversial subject for a period of several decades. The disagreements were not only of mathematical nature but also a matter of definitions. During the past ten years most of the discrepancies and misunderstandings have been cleared up. We now have a rather good understanding of the fields on interfaces, although there are still basic questions open which require further study.

Someone who is not familiar with this subject may ask about the peculiarity of electromagnetic wave propagation on interfaces and in what respect it differs from optical wave propagation. The answer to this question is that in any optical experiment, source and point of observation are far away from the interface. Under these circumstances the wave field is adequately described by the well-known optical refraction and diffraction laws. Electromagnetic sources can be placed arbitrarily close to an interface and the field can be observed in proximity to it. For such conditions, optical laws no longer prevail. The spatial distribution of the energy delivered by the source is greatly affected by the presence of the interface, and part of the energy has the tendency to propagate along the interface.

Figs. 1 and 2 illustrate two basic examples of sources placed close to an interface. Fig. 1 shows an electric dipole above a plane interface between a nonconductive and a conductive medium. Fig. 2 shows the excitation of a wire by a magnetic ring current. In this case the interface is cylindrical. The magnetic ring current is the equivalent of a voltage source inserted into the wire. These two examples, though very simple from the physical point of view, have been the major subject of the mentioned controversy.

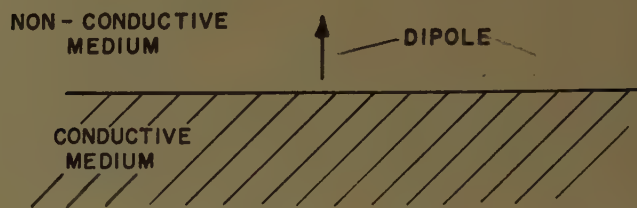


Fig. 1.

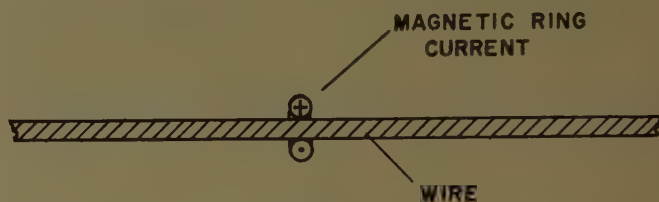


Fig. 2.

II. HISTORY

Before we enter into a more detailed discussion on wave propagation along interfaces, we will present a short historical review.

Hertz [1], in 1888, was the first to investigate the propagation along a conductive wire. He expected the phase velocity of the waves to be the same as in free space. However, he was not able to prove this, either experimentally or theoretically. The experimental difficulties were caused by the fact that the field was not sufficiently confined to the wire. Reflections from the walls of the room interfered with his measurements. His theoretical approach failed because he treated the wire as infinitely thin. Thus, he could not establish approximate boundary conditions.

In 1890, Lecher [2] introduced the two-wire system which proved to have much better guiding properties than a single wire. The field is more confined to the proximity of the wires and is not so much affected by the surroundings. Lecher's two-wire system became an important device for measuring wavelengths. The two-wire line can be considered as the prototype of open waveguides.

J. J. Thomson [3] first treated the coaxial line in 1893. He introduced correct boundary conditions, but, in trying to solve the resulting transcendental equation, made severe simplifications. His end result does not show any effect of the wire material on phase velocity and attenuation of the waves if the radius of the outer conductor is made infinitely large. The phase velocity becomes equal to the free-space velocity. He, therefore, assumes this velocity in his treatment of Hertz's wire waves.

† U. S. Army Signal Research and Development Laboratory, Ft. Monmouth, N. J.

In 1899, Sommerfeld [4] found the first rigorous solution for the boundary value problem on a single wire. He assumed that there was an axially symmetrical, cylindrical, transverse, magnetic wave traveling along the wire, and was able to satisfy the boundary conditions. The phase velocity of this wave depends on the conductivity of the wire material and is smaller than the free-space velocity. Sommerfeld considered this problem as an idealized one, stating in his paper that he would neglect the effect of the "return current" in his analysis. He also suggested that this treatment may be considered as a supplement to Thomson's investigations. Sommerfeld, as well as Thomson, thought that their results were applicable to Hertz's wire wave, assuming that in Hertz's experiments part of the current returned through the walls of the room. Today we know that Hertz's experimental set-up was not appropriate to excite a wave of the kind investigated by Thomson and Sommerfeld.

In 1900, Mie [5] succeeded in solving the boundary value problem for a two-wire line consisting of identical wires with finite conductivity. He obtained two solutions: one describing the Lecher wave, and the other a wave with in-phase currents in both wires.

Plane waves guided by a plane interface between an insulator and a good conductor were first investigated by Uller [6] in 1903. Zenneck [7], in 1907, recognized the bearing of these studies on the propagation of radio waves along the earth. He investigated the case where one half-space is a pure dielectric and the other half-space is a dielectric which is more or less conductive. Zenneck was quite aware of the limitations of his results. However, he succeeded in explaining several characteristic phenomena which had been observed in radio wave transmission.

Harms [8], in 1907, extended Sommerfeld's work to a wire which is surrounded by a dielectric sheath. His work had been initiated by an observation of Slaby [9] who found that the resonance length of an antenna built of a heavily insulated wire is shorter than that of an antenna made of bare wire. However, the conditions in Slaby's experiments were not adequate to excite a wave of the type investigated by Harms.

Hondros [10] in 1909, extended Sommerfeld's work in another direction. Sommerfeld considered only one possible solution of his boundary value problem. In today's terminology, his wave is called the "fundamental mode." Hondros investigated the higher modes, which have extremely high attenuation since most of the energy is inside the conductor. As a continuation of this work, Hondros and Debye [11] investigated the higher mode solutions for a nonconductive dielectric cylinder in 1910. All these modes exist only above a certain cutoff frequency which is determined by the diameter of the cylinder and the dielectric constant. The field of the fundamental mode, the so-called dipole mode, which has no cutoff frequency was derived much later by Schelkunoff [37].

In 1909, a major step forward in the theory of wave propagation along interfaces was made by Sommerfeld [13] when he solved the problem of a dipole radiating above the plane interface of a nonconductive and a conductive medium. Here, for the first time, the source was taken into account and a complete solution of a boundary value problem was obtained. All the previously discussed solutions were incomplete, since they only considered boundary conditions at the interfaces but not at the source. The question as to whether these solutions could be physically realized remained unanswered. Sommerfeld obtained his solution in the form of a complex integral. The integrand of this integral comprises a pole, and the field associated with this pole describes a cylindrical Zenneck wave. Thus, Sommerfeld divided the total field into two parts: a surface wave of the type investigated by Zenneck, and a space wave.

In 1919 Weyl [14] treated the same problem in a somewhat different manner. He arrived at the conclusion that Sommerfeld's separation of the field into a surface wave and a space wave was arbitrary and that there is no Zenneck wave excited. From this time on, there appeared a large number of papers pro and con the Zenneck wave. Until a few years ago when various kinds of surface waves became of practical interest, the majority of investigators decided against the Zenneck wave.

Wave propagation along a spherical interface was first treated by Watson [15], in 1918, with the assumption that the sphere is a perfect conductor, and by Van der Pol and Bremmer [16], in 1937, for a finitely conducting sphere. The latter case is of fundamental importance for the propagation of radio waves along the earth.

A rigorous approach to the excitation problem of a wire by a concentrated power source was first made by Hallén [17], in 1938. His results and also those of King [18], Schelkunoff [19], and of others who studied the same problem, did not show any evidence of Sommerfeld's wire wave. Since there was no experimental evidence of Sommerfeld's wave, and since their results were in agreement with experiments, the opinion of most experts was that this wave, like the Zenneck wave, is unreal. Both waves are closely related in that the former transforms into the latter if the radius of the wire approaches infinity.

It is impossible to quote in this short historical review all the *important* contributions to the problem of wave propagation along interfaces. I therefore mentioned only those which I consider as the "milestones" on the road leading to our present understanding of the waves along boundaries.

III. SURFACE WAVES AND RADIATING WAVES

The quoted literature demonstrates quite clearly that no attention had been given to the structure of the source. As far as sources had been included in the theory,

they were of the most elementary types, namely dipoles or concentrated voltage sources. No doubt, any source can be thought of as composed of distributed dipoles and the corresponding field as the superposition of the fields excited by the individual dipoles. However, this procedure yields little insight into the special properties of the fields on interfaces.

Although it is obvious that different sources yield different fields, we can distinguish between two basically different kinds of waves, namely "truly guided waves," like the Lecher wave or Sommerfeld's wire wave, where the entire energy is confined to the proximity of the interface, and "partially guided waves," where the energy propagated along the interface decreases because of radiation into space. In both cases, there may be losses caused by absorption within the media. The first kind of wave is called, in this paper, "surface wave" and the second kind "radiating wave."

If a boundary value problem has surface wave solutions, it is a matter of the structure of the source whether one or the other kind of wave is prevailing.

It is obvious that any separation of a complex field into components is more or less arbitrary in that there are various points of view under which a separation may be performed. The separation discussed in this paper is very useful in cases of open waveguides like two-wire lines, single conductor surface wave transmission lines, and dielectric rod waveguides, where simple sources can be devised which substantially excite only a surface wave. In other cases, like the wave propagation along the earth, this kind of separation is of little practical value since present-day antennas excite fields which differ greatly from a Zenneck wave. A separation as proposed by Norton [38] is much more appropriate here.

Since the term, surface waves, is now used for various kinds of fields near interfaces, it shall be emphasized that the surface waves considered in this paper are non-radiating waves as mathematically defined in Section IV.

The following examples may demonstrate the difference between surface waves and radiating waves from the physical point of view.

Consider first a plane interface between a dielectric half-space and an ideal conductor as in Fig. 3. The source may be a vertical electric dipole located at, or above, the interface. The field within the dielectric half-space can be considered as the superposition of the direct field of the source and the field of the image of the source. Although the resulting field near the interface is a wave propagating parallel to the interface, it is usually not called a guided wave. If we wish to classify it in one of the two categories of waves, we have to classify it as a radiating wave.

If the interface is curved, as shown in Fig. 4, the field along the interface beyond line of sight is a diffraction field. But this diffraction field differs from that which is observed behind a diffracting edge in that there is a continuous or differential diffraction taking place. In this case, we may speak of a guided wave because the

field near the interface depends primarily on the curvature of the interface. It is a radiating wave since part of the energy which passes through an area S above the interface (see Fig. 4) is radiated into space.

As a third case, we consider an infinitely long wire of infinite conductivity excited by a magnetic ring current (see Fig. 5). Here the guiding effect is much more pronounced than in the preceding case. The energy travels substantially along the wire, but the field strength at the surface of the wire decreases with increasing distance from the source and approaches zero at infinity. In other words, the wave which is excited on the wire is attenuated because of radiation. This is a clear-cut case of a partially guided or radiating wave.

Now we assume that there is a second wire parallel to the first one, also having infinite conductivity (Fig. 6). This case is basically different from the preceding one in that the field strength near the wires approaches a finite value at infinity. In other words, the coupling of the second wire prevents the energy from being completely radiated. This appears obvious if we consider the problem *a priori* as an excitation problem of an ideal two-wire line. It is not at all obvious if we consider it from the viewpoint of a coupling problem. The total field can thus be divided into two components: a surface wave, namely a Lecher wave, and a radiating wave. Either of

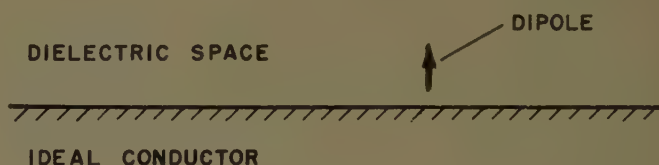


Fig. 3.

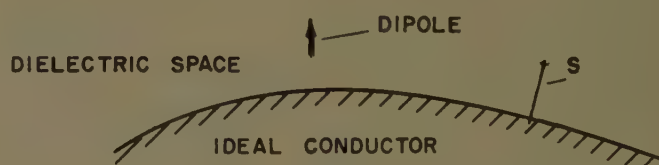


Fig. 4.

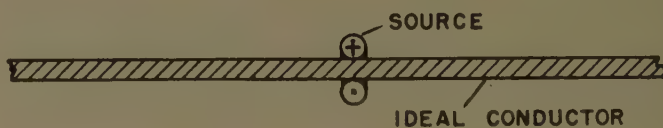


Fig. 5.

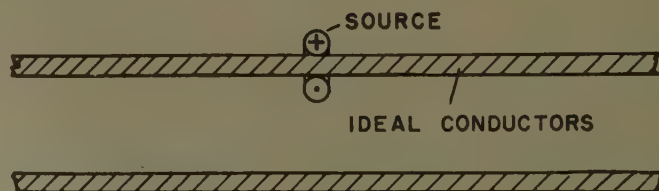


Fig. 6.

these two waves satisfies the boundary conditions at the wires. Both together are required to satisfy the boundary conditions at the source. The power delivered by the source is the sum of powers propagated by the two waves.

Thus far we have considered only lossless structures. Now, we assume that the wires of Fig. 6 have finite conductivity. To avoid unnecessary complications we furthermore assume that the wires are alike and are excited by two identical, but oppositely directed, sources as indicated in Fig. 7. In other words, we consider symmetrical excitation of an *actual* two-wire line. The only remaining idealizations are that the line is infinitely long and perfectly straight. By virtue of the push-pull excitation, the radiation is reduced as compared to the preceding case, but there is still some radiation present. The existence of a surface wave and its amplitude can only be deduced from the exponential decrease of the field along the wires. The amplitude at infinity is no longer finite because of the losses in the wires. There is another basic difference between this case and the preceding one. The power delivered by the sources can no longer be expressed by the sum of the powers propagated by the surface wave and the radiating wave. But, as before, each of the waves satisfies the boundary conditions at the wire and both are required to satisfy the conditions at the sources.

Now we replace the two sources by a single magnetic dipole as indicated in Fig. 8, still maintaining push-pull excitation. In this case, the coupling between source and line is greatly reduced, particularly if the dipole is moved farther away from the line. The direct radiation of the dipole becomes the prevailing field, and the field near the line may not have any similarity to a Lecher wave.

Under these circumstances, the question appears justified as to whether there is still a surface wave present. We cannot answer this question by measuring the field distribution. Strictly speaking, we cannot even do this in the preceding case with rather ideal excitation conditions for a two-wire line because the assumed ex-

ponential field decrease is not exact. Since there is only a gradual difference between the two cases, it would be illogical to assume the presence of a surface wave in one case and not in the other. Therefore, we have to accept the fact that the surface wave is still present even if we may not be able to detect or separate it.

IV. DEFINITION OF SURFACE WAVES FROM THE MATHEMATICAL POINT OF VIEW

Surface waves, from the mathematical point of view, are particular solutions of the wave equation, which exist in certain boundary value problems where homogeneous or stratified media are separated by plane or cylindrical interfaces. These solutions describe cylindrical waves whose propagation vector is parallel to the interface. In the case of cylindrical interfaces, the waves propagate parallel to the cylinder axis, and in the case of plane interfaces, the propagation is radial. Plane waves may be considered as asymptotic solutions of the latter case.

Only such solutions have physical meaning which can be realized experimentally. The only requirement for this is that the power propagated by the wave is finite. In all solutions where this condition is satisfied, the field decreases exponentially at large distances from the interface. An exception is the ideal two-wire line whose field decreases with the square of the distance. But, since all materials have losses, we can disregard this case.

An ideal source for a surface wave on a cylindrical structure is a layer of electric and magnetic dipoles appropriately distributed over a plane perpendicular to the cylinder axis. The dipole densities P and M , respectively, are given by the relations,

$$j\omega P = -(H_s \times n), \quad j\omega M = (E_s \times n), \quad (1)$$

where E_s and H_s are the field vectors of the surface wave at the plane, and n is the unit vector parallel to the axis.

Surface waves on plane interfaces require a cylindrical layer of dipoles, the axis of this cylinder being vertical to the interface. The dipole densities are also given by above relations. In this case, n is radially directed.

Although the field of a surface wave decays exponentially in directions perpendicular to the interface, it still extends to infinity. The required dipole layer can therefore be simulated only within a finite area. For this reason, a surface wave never describes the complete field in an actual excitation problem. However, if a finite volume of the total space is considered, the deviation of the actual field from that of the surface wave can be made arbitrarily small if the source is made large enough. But far away from the source, its limitation in size always becomes apparent because the asymptotic field is never determined by the surface wave but by the radiating wave, except if there are no losses.

It shall be emphasized that the phase velocity of a surface wave has no bearing on its realizability. The existence of the Zenneck wave has sometimes been de-

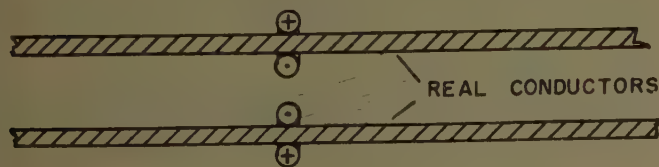


Fig. 7.

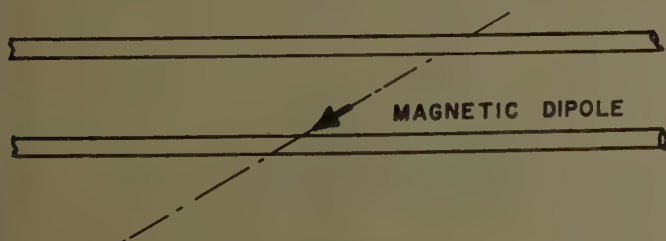


Fig. 8.

nied because its phase velocity is greater than that of light. The phase velocity of a surface wave can be greater, equal, or smaller than c . Barlow and Karbowskiak [30] demonstrated on a cylindrical dielectric rod with appropriate losses an axial symmetrical surface wave which has a phase velocity greater than c .

V. STRUCTURES WHICH HAVE SURFACE WAVE SOLUTIONS

A summary of surface wave supporting structures including some literature quotations is given in Table I.

Surface wave solutions are also obtained for corrugated metal surfaces or periodic structures like a column of alternating dielectric and metal disks. When the corrugations are small or the periodicity is short compared to the wavelength one can introduce average boundary conditions. In this manner, the surface wave is characterized by a well defined propagation constant like any other surface wave. The correct boundary conditions, however, require the wave to consist of a bundle of waves with a spectrum of propagation constants. To my knowledge, that such a bundle has no radiating components has not yet been proved.

VI. RELATION BETWEEN SURFACE WAVES AND RADIATING WAVES

The separation of the field on an interface into a surface wave and a radiating wave may still appear arbitrary or at least unsatisfactory if we have no method for separating the two field components. Even if we do not insist on an experimental separation, we would like to have at least a mathematical relation which enables us to determine the amplitude of the surface wave if the total field is known. We cannot rely on the asymptotic behavior of the field since every physical structure has losses and since there may not be any range in which the field is substantially determined by the surface wave.

In order to perform the desired separation, it is necessary to have not only a definition for a surface wave but also a definition for a radiating wave. Such a definition is obtained from the following consideration. Consider an ideal antenna which excites only a surface wave. We postulate that such an antenna, used as a receiving antenna, should not respond to a radiating wave. The mathematical formulation of this condition leads to the following orthogonality relations between surface waves and radiating waves.

$$\int_S (E_S \times H_R) n dS = \int_S (E_R \times H_S) n dS = 0, \quad (2)$$

where E_S , H_S and E_R , H_R are the field vectors of the surface wave and the radiating wave. The integration is extended over an equiphase surface of the surface wave. These relations have been derived by means of the reciprocity theorem and are valid for the waves on cylindrical interfaces as well as for those on plane interfaces. They can be easily verified on lossless structures where the surface waves determine the asymptotic field [32].

TABLE I

| A. PLANE INTERFACES | | |
|---------------------------|--|------------------|
| a) | | [7] [13] |
| b) | | [20]-[22] |
| * c) | | |
| B. CYLINDRICAL INTERFACES | | |
| a) | | [4] |
| b) | | [8] [23] [24] |
| c) | | [5] |
| d) | | [25] |
| e) | | [26] |
| f) | | [27]-[28] |
| g) | | [29] |
| h) | | [11], [37], [12] |
| i) | | [30] |
| C. UNISOTROPIC CASE | | |
| | | [31] |

* There are two surface waves of different phase velocity.

The same orthogonality relations hold for closed waveguides between the various wave modes or between any mode and the remainder field. They also apply to waveguides with losses where orthogonality with respect to power is, in general, no longer valid.

Above orthogonality relations uniquely define a radiating wave which contains no surface wave component. If there are more surface wave modes in existence, the relations must be satisfied for every mode. In this case, the surface wave modes are also orthogonal to each other [33].

VII. EXCITATION OF SURFACE WAVES

By use of the above orthogonality relations, the amplitude of a surface wave can be easily determined by means of the reciprocity theorem. For instance, for a surface wave on a cylindrical structure which is excited by an electric or magnetic dipole, one obtains the amplitude [32]

$$\alpha = -\frac{j\omega(PE_s)}{2N} \text{ or } \frac{j\omega(MH_s)}{2N}$$

with

$$N = \int_S (E_s \times H_s) n dS, \quad (3)$$

where E_s , H_s are the field vectors of a surface wave of unit amplitude (without the propagation factor).

In Sommerfeld's problem on the radiation of a vertical dipole above the earth, the amplitude of the Zenneck wave is [34]

$$\alpha = -\frac{j}{2} \frac{\omega P}{A_0} \sqrt{\frac{\mu'}{\epsilon}} \frac{\gamma}{k} \frac{h^2 h'^2}{h'^2 - h^2} e^{-i h z_0}$$

with

$$h' = h\epsilon'/\epsilon, \quad k^2 = \omega^2\epsilon\mu = h^2 + \gamma^2, \quad (4a)$$

where A_0 is the unit amplitude of the radial electric field component E_r of a unit Zenneck wave

$$(E_r = A_0 H_1(\gamma r) e^{-j(\omega t - h z)})$$

and z_0 is the height of the dipole above ground. ϵ , ϵ' and μ , μ' are dielectric constant and permeability of air and ground respectively.

If the dipole is below the surface of the earth ($z_0 < 0$), the amplitude of the Zenneck wave becomes

$$\alpha = -\frac{j}{2} \frac{\omega P}{A_0} \sqrt{\frac{\mu'}{\epsilon'}} \frac{\gamma}{k'} \frac{h^2 h'^2}{h'^2 - h^2} e^{-i h' z_0}$$

with

$$k'^2 = \omega^2\epsilon'\mu' = h'^2 + \gamma^2. \quad (4b)$$

An ideal source for the Zenneck wave is a cylindrical layer of dipoles whose distribution is given by (1). If

the diameter of the cylinder is assumed to be very small, one can easily verify that (4a) and (4b) yield the correct amplitude of the Zenneck wave.

It shall be emphasized again that if there are losses involved as in the case of the Zenneck wave, the total power is not the sum of the power transmitted by the surface wave and the power transmitted by the radiating wave.

The excitation of a surface wave is always connected with simultaneous excitation of a radiating wave. This is true for the Lecher wave as well as for any other surface wave. However, the radiated energy can be made very small if the source is appropriately designed. In order to excite a surface wave efficiently, it is, in general, necessary to build up a field which matches that of the surface wave. Efficient excitation is a matter of "field matching" and *not*, as one can find occasionally in the literature, a matter of "impedance" matching. Impedance is an ambiguous quantity in that it is not uniquely defined.

In certain cases, efficient surface wave excitation can be achieved without field matching. A well known case is the two-wire line which can be excited by a concentrated source, with very little radiation if the distance of the wires is small compared to the wavelength. Two other interesting cases have been investigated by Cullen [35] and Brown and Stachera [36]. Cullen studied the excitation of plane surface waves on a dielectric covered metal plane by means of a line or slot source. If the source has a certain height above the surface, a launching efficiency of about 95 per cent should be obtainable. Brown and Stachera investigated the excitation of a dielectric covered wire by an annular slot in a metal wall terminating the wire. They also obtained a theoretical efficiency of 95 per cent. The measured efficiency was about 65 per cent. In both these cases, the supplementary field which is required to meet the boundary conditions at the source is primarily a reactive field. The excitation is therefore substantially frequency dependent.

Another way to excite surface waves on cylindrical structures without field matching is to build a resonator consisting of a section of a surface waveguide terminated at both ends by large metal plates. Such resonators are frequently called "open cavities" or "parallel plate cavities" and are used to measure transmission losses of surface waveguides. If the line section is made many wavelengths long, the resonance of the radiating wave is practically not detectable.

VIII. SOME REMARKS ABOUT THE ZENNECK WAVE

The existence of the Zenneck wave has been disputed so much in the literature that it may be excusable if a few more remarks are added to this subject. The dispute originated in the attempt to extract from the mathematical solution of a problem more answers than there were questions when the problem was formulated mathematically.

If one formulates the problem of a dipole radiating above a plane interface, one can only expect a solution which describes the *total* field of the dipole, no matter what mathematical method one uses. There is no obligation from the physical point of view to use complex variables for solving the problem. The fact that the problem is more accessible to a rigorous treatment, when solved in the complex plane, is a purely mathematical matter and one cannot expect this method to yield more information than a treatment with only real quantities. If Sommerfeld had solved the problem by use of real quantities only, it is unlikely that the question as to a surface wave and a space wave would ever have arisen (except perhaps now) since the actual field of a dipole at the interface differs substantially from that of the surface wave.

The fact that the answer to a given problem can be written in terms of a complex integral whose integrand comprises a pole, does not *a priori* mean that this pole has physical meaning. Even if this is the case, there is still no obligation that the integration has to be performed in such a manner that the pole is included.

In order to separate the field into two components, it is necessary to define both these components. Defining only one, namely the surface wave, is not enough. However, I believe that the orthogonality relations quoted in this paper should fill this gap and present a satisfactory basis for the separation.

BIBLIOGRAPHY

- [1] H. Hertz, "Gesammelte Werke," J. A. Barth, Leipzig, Germany, vol. II; 1894.
- [2] E. Lecher, "Eine Studie über elektrische Resonanzerscheinungen," *Wied. Ann. Phys. u. Chemie*, vol. 41, p. 850; December, 1890.
- [3] J. J. Thomson, "Notes on Recent Researches in Electricity and Magnetism," Oxford University Press, London, Eng., pp. 261-268; 1893.
- [4] A. Sommerfeld, "Ueber die Fortpflanzung elektrodynamischer Wellen laengs eines Drahtes," *Wied. Ann. Phys. u. Chemie*, vol. 67, pp. 233-290; February, 1899.
- [5] G. Mie, "Elektrische Wellen an zwei parallelen Prähten," *Ann. der Phys.*, vol. 2, p. 201; July, 1900.
- [6] K. Uller, Ph.D. dissertation, University Rostock, Germany; 1903.
- [7] J. Zenneck, "Ueber die Fortpflanzung ebener elektromagnetischer Wellen laengs einer ebenen Leiterflaeche," *Ann. der Phys.*, vol. 23, pp. 846-866; September, 1907.
- [8] F. Harms, "Elektromagnetische Wellen an einem Draht mit isolierender zylindrischer Huelle," *Ann. der Phys.*, vol. 23, pp. 44-60; May, 1907.
- [9] A. Slaby, "Die Abstimmung funkentelegraphischen Sender," *Elektrotech. Z.*, vol. 26, pp. 1003-1007; November, 1905.
- [10] D. Hondros, "Ueber elektromagnetische Drahtwellen," *Ann. der Phys.*, vol. 30, p. 905; December, 1909.
- [11] D. Hondros and P. Debye, "Elektromagnetische Wellen an dielektrischen Draehten," *Ann. der Phys.*, vol. 32, pp. 465-476; June, 1910.
- [12] W. M. Elsasser, "Attenuation in a dielectric rod," *J. Appl. Phys.*, vol. 20, pp. 1193-1196; December, 1949.
- [13] C. M. Chandler, "An investigation of a dielectric rod as a waveguide," *J. Appl. Phys.*, vol. 20, pp. 1188-1192; December, 1949.
- [14] A. Sommerfeld, "Ueber die Ausbreitung der Wellen in der drahtlosen Telegraphie," *Ann. der Phys.*, vol. 28, pp. 665-736; March, 1909.
- [15] H. Weyl, "Die Ausbreitung elektromagnetischer Wellen ueber einem ebenem Leiter," *Ann. der Phys.*, vol. 60, pp. 481-500; November, 1919.
- [16] G. N. Watson, "The diffraction of electric waves by the earth," *Proc. Roy. Soc. (London)*, vol. 95A, pp. 83-99; October, 1919.
- [17] B. van der Pol and H. Bremmer, "The diffraction of electromagnetic waves from an electrical point source round a finitely conducting sphere," *Phil. Mag.*, vol. 24, pp. 141-176, 825-864; July and November, 1937.
- [18] E. Hallén, "Theoretical investigations into the transmitting and receiving qualities of antennae," *Nova Acta Regiae Soc. Sci. Upsaliensis*, vol. 11, pp. 1-44; November, 1938.
- [19] R. W. P. King and D. Middleton, "The thin cylindrical antenna," *J. Appl. Phys.*, vol. 17, pp. 273-284; April, 1946.
- [20] S. A. Schelkunoff, "Electromagnetic Waves," D. Van Nostrand Co., Inc., New York, N. Y., pp. 441-471; 1943.
- [21] S. C. Attwood, "Surface wave propagation over a coated plane conductor," *J. Appl. Phys.*, vol. 22, pp. 504-509; April, 1951.
- [22] Y. T. Lo, "Electromagnetic field of a dipole source above a grounded slab," *J. Appl. Phys.*, vol. 25, pp. 733-740; June, 1954.
- [23] D. B. Brick, "The radiation of a Hertzian dipole above a coated conductor," *Proc. IEE*, Pt. C, vol. 102, pp. 104-121; March, 1955.
- [24] J. R. Wait, "Excitation of surface waves on conducting, stratified, and dielectric clad surfaces," *Natl. Bur. Standards J. Research*, vol. 59, pp. 365-377; December, 1957.
- [25] G. Goubau, "Surface waves and their applications to transmission lines," *J. Appl. Phys.*, vol. 21, pp. 1119-1128; November, 1950.
- [26] T. E. Roberts, "Theory of single-wire transmission line," *J. Appl. Phys.*, vol. 24, pp. 57-67; January, 1953.
- [27] G. Goubau, "Open wire lines," *IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES*, vol. MTT-4, pp. 197-200; October, 1956.
- [28] A. Meyerhoff, "Interaction between surface-wave transmission lines," *Proc. IRE*, vol. 40, pp. 1061-1065; September, 1952.
- [29] F. Pollaczek, *Elek. Nachrichten Tech.*, vol. 3, pp. 339-359; September, 1926.
- [30] H. Kikkuchi, "Wave propagation along an infinite wire above ground in the high-frequency region," *Bull. Electrotech. Lab. Tokyo*, vol. 21, pp. 59-61; 1957.
- [31] D. D. Grieg and H. F. Engleman, "Microstrip, a new transmission technique for the kilomegacycle range," *PROC. IRE*, vol. 40, pp. 1644-1650; December, 1952.
- [32] H. E. M. Barlow and A. E. Karbowiak, "An experimental investigation of axial cylindrical surface waves by capacitive surfaces," *Proc. IEE*, Pt. B, vol. 102, pp. 313-322; May, 1955.
- [33] R. L. Pease, "On the propagation of surface waves over an infinite grounded ferrite slab," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-6, pp. 13-20; January, 1958.
- [34] G. Goubau, "On the excitation of surface waves," *PROC. IRE*, vol. 40, pp. 865-868; July, 1952.
- [35] R. Adler, "Waves on inhomogeneous cylindrical structures," *Proc. IRE*, vol. 40, pp. 339-348; March, 1952.
- [36] G. Goubau, "Ueber die Zenneck'sche Bodenwelle," *Z. angew. Phys.*, vol. 3, pp. 103-107; 1951.
- [37] A. L. Cullen, "The excitation of plane surface waves," *Proc. IEE*, pt. IV, vol. 101, pp. 225-234; August, 1954.
- [38] J. Brown and H. S. Stachera, "Annular-slot launchers for single-conductor transmission lines," *IEE Mono. #2836*; January, 1959.
- [39] S. A. Schelkunoff, "Electromagnetic Waves," D. Van Nostrand Co., Inc., New York, N. Y., pp. 425-430; 1943.
- [40] K. A. Norton, "The physical reality of space and surface waves in the radiation-field of antennas," *Proc. IRE*, vol. 25, pp. 1192-1202; September, 1937.

Surface Waves Supported by Cylindrical Surfaces

H. M. BARLOW†

Summary—In approaching a discussion of this matter, it is most important to ensure general agreement about the particular features that characterize the so-called surface wave. An attempt has been made to formulate in simple and unambiguous terms a definition of this form of wave, distinguishing it from the various other waves known to be associated with an interface between two different media. It is maintained that there are really two closely allied aspects of the problem of surface wave propagation, one of which is more particularly concerned with launching the wave from a given aerial over the surface and the other with the capabilities of that surface in supporting such a wave. The present paper deals with the latter consideration in relation to surfaces slightly curved in the direction of the propagation of the wave.

Recognizing the evanescent character of the surface wave field distribution over the equiphase surfaces and the important part played by the inclination of these surfaces with the normal to the interface when power is transferred across it, a method is discussed of calculating the radiation which arises when a wave of this kind circulates around a highly reactive supporting surface of cylindrical form. It is concluded that when the surface has a finite loss there will be a particular radius of curvature for which no power is transferred across the interface.

I. INTRODUCTION

SEVERAL attempts have been made [1] to define in unambiguous terms what is now widely recognized as a surface wave. As the name implies, such a wave must be associated with an interface between two different media, or, in other words, there must be a recognizable surface on which to support the wave and it is therefore of the kind we call a guided wave. Bearing this in mind, we might perhaps suitably describe a surface wave by saying that it transmits no power away from the supporting surface except that which is required to supply losses associated with the wave in the surrounding media. As a corollary to this statement, we should define radiation as power transmitted by a field not identifiable as a characteristic mode of the waveguide structure.

Most of the common forms of surface wave are *E* modes which have a component of electric field in the direction of propagation, but there are also *H* modes and hybrid *EH* modes supported by cylindrical guides. To satisfy the surface wave criterion that the only power directed away from the supporting surface must be absorbed by losses arising from the wave in the media on each side of the boundary, we can offer an equivalent specification simply requiring that either the electric or magnetic field component which is tangential to the supporting surface have an evanescent distribution over the corresponding equiphase surface. In formulating this specification as a possible alternative means of defining a surface wave, due regard has been given to

the fact that the various field components of any such wave may have different equiphase surfaces. All that is really necessary to meet the requirements is that, at any particular point, one of the field components which is tangential to the supporting surface have the evanescent structure which is the essential characteristic of a surface wave.

When the supporting surface is straight in the direction of propagation of the wave, it is generally a comparatively simple matter to ascertain whether there is a solution to Maxwell's equations representing a field distribution of the surface wave form. If there is such a solution, it seems justifiable to conclude that the corresponding surface wave can be supported by the surface in question and, moreover, that a radiating aperture of infinite area is required to guarantee that the wave can be launched. We know from experience in other cases that a particular wave mode can often be quite strongly excited even when the radiating aperture is of a much smaller area than the cross section of the wave which it is required to launch. This is particularly true for a resonant structure like a hollow metal waveguide where there is a pronounced predisposition to the formation of certain defined wave modes. It is less true of surface waves, but there is, nevertheless, the same kind of tendency because these waves are also characteristic modes of the guiding structure. Much work has already been done [2], both mathematically and experimentally, to establish the degree to which surface waves can be excited over a given surface when different radiating systems are used. This is, of course, a very vital aspect of the surface wave problem because it is of little use to know that a particular surface is capable of supporting a surface wave if there is no known radiator of manageable dimensions capable of exciting it with reasonable efficiency. Nevertheless, it must be emphasized that there are these two aspects of the problem, namely launching and support, which can quite legitimately be independently discussed.

When the supporting surface is curved in the direction of propagation of the wave, another important factor is introduced because in these circumstances there is always a tendency for radiation to be set up. If we imagine a surface surrounded by air and purely reactive; *i.e.*, lossless, then, no matter how small the curvature, it is impossible to avoid radiation. However, for highly lossy surfaces which must also be reactive, it may be possible to balance the power which would be radiated from the corresponding ideal surface against some of the power required to supply the losses in the actual surface, and thus eliminate transfer of energy across the interface. We will first discuss a possible physical picture

† University College, London, England.

of the mechanism of radiation when a surface wave is supported by a cylindrical surface, and then proceed to calculate the radial and circumferential power.

II. THE AZIMUTHAL SURFACE WAVE

To provide for the support of a close approach to the Zenneck form of inhomogeneous plane wave outside a cylindrical surface of very slight curvature (see Fig. 1),

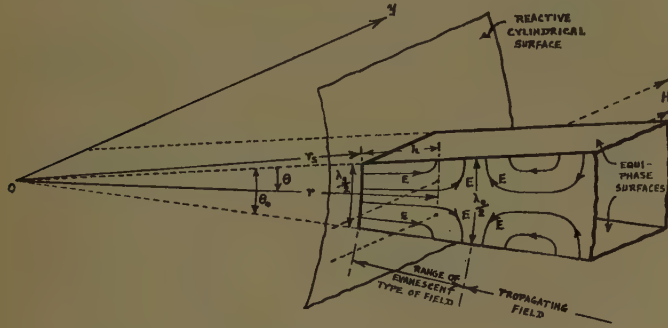


Fig. 1—Coordinate system and field pattern outside supporting surface.

it is essential that the surface be highly reactive, storing the energy of a trapped wave from which leakage occurs as the wave progresses. Elliott [3] has examined the behavior of this kind of azimuthal surface wave when it is associated with dielectric-coated and corrugated cylindrical metal surfaces. His analysis leads to surprisingly small values of circumferential attenuation, and although there is reasonable correlation between some values of the propagation coefficients for the two types of supporting surface, curious inconsistencies seem to arise when these coefficients are small. No doubt the difficulty lies very largely in the approximations that have to be made in trying to get a manageable solution without departing too far from the very exacting conditions of any practical application, but there also seems to be a lack of recognition of the part played by any field reflected from outer space when the cylindrical symmetry is disturbed. The author has discussed [4] this aspect of the problem of the circulating surface wave in an attempt to trace the physical processes from which radiation arises in a practical case and to estimate its magnitude under prescribed conditions. The aim of the present paper is to develop the argument further, by taking into account the effect of a field reflected from outside the surface and by progressing to a calculation of both the radial and circumferential components of power.

A. Field Equations

Consider the cylindrical coordinate system shown in Fig. 1, with field components E_r , E_θ and H_y , there being no change in the y direction. Assuming that there is a sinusoidal time variation $e^{j\omega t}$ and a homogeneous medium of permittivity ϵ , permeability μ and conductivity σ , we find that

$$\frac{\partial^2 H_y}{\partial r^2} + \frac{1}{r^2} \frac{\partial^2 H_y}{\partial \theta^2} + \frac{1}{r} \frac{\partial H_y}{\partial r} + T^2 H_y = 0, \quad (1)$$

where the field component H_y is now dependent upon space coordinates only, and

$$T^2 = -j\omega\mu(\sigma + j\omega\epsilon). \quad (2)$$

Hence,

$$H_y = Af_1(\theta)f_2(r), \quad (3)$$

and for a circulating wave travelling in the $+\theta$ direction, we have

$$f_1(\theta) = e^{-j\nu\theta}. \quad (4)$$

From (1), (3), and (4) we therefore get

$$\frac{\partial^2 f_2}{\partial r^2} + \frac{1}{r} \frac{\partial f_2}{\partial r} + \left(T^2 - \frac{\nu^2}{r^2}\right) f_2 = 0 \quad (5)$$

to which the general solution¹ applicable here is

$$f_2 = B_1 H_{\nu}^{(2)}(Tr) + B_2 H_{-\nu}^{(1)}(Tr), \quad (6)$$

where both ν and T are, as a rule, complex.

Since we are concerned with a surface wave, the field distribution in the neighborhood of the surface must closely approach an evanescent structure. Thus, if the radius of the surface is r_0 , we must have $|Tr_0| < |\nu|$ by only a few per cent and $|\nu|$ of the order of 100 upwards.

In these rather difficult circumstances, we can apply Liouville's approximations to the Bessel functions over the limited range of interest to show that $B_1 H_{\nu}^{(2)}(Tr)$ represents the outgoing or "incident" field which decreases as r increases, while $B_2 H_{-\nu}^{(1)}(Tr)$ is the corresponding "reflected" field. The presence of any such reflected field implies a physical discontinuity disturbing the progress of the incident field. The field components can be written as

$$H_y = H_y^+ + H_y^- = e^{-j\nu\theta} [C_1 H_{\nu}^{(2)}(Tr) + C_2 H_{-\nu}^{(1)}(Tr)], \quad (7)$$

$$\begin{aligned} E_\theta &= E_\theta^+ + E_\theta^- \\ &= - \left[\frac{e^{-j\nu\theta}}{\sigma + j\omega\epsilon} \right] \left[C_1 \left\{ \frac{\nu}{r} H_{\nu}^{(2)}(Tr) - T H_{\nu+1}^{(2)}(Tr) \right\} \right. \\ &\quad \left. + C_2 \left\{ -\frac{\nu}{r} H_{-\nu}^{(1)}(Tr) - T H_{-\nu+1}^{(1)}(Tr) \right\} \right], \end{aligned} \quad (8)$$

and

$$\begin{aligned} E_r &= E_r^+ + E_r^- = - \left[\frac{j\nu e^{-j\nu\theta}}{r(\sigma + j\omega\epsilon)} \right] \\ &\quad \cdot [C_1 H_{\nu}^{(2)}(Tr) + C_2 H_{-\nu}^{(1)}(Tr)] \end{aligned} \quad (9)$$

We shall see later that the field in the vicinity of the curved surface can never be purely evanescent, and that even when there is no reflected field some radial flow of energy must occur. Thus, with a lossless medium out-

¹ Since $H_{-\nu}^{(1)}(Tr) = e^{j\nu\theta} H_{\nu}^{(1)}(Tr)$, there is only a constant factor between these two forms and the first is preferred in this solution.

side the surface and a corresponding real value of T , the quantity ν is still complex, providing for the appropriate circumferential attenuation of the wave. It will be observed that ν and the complex coefficients C can be determined from the matching at $r=r_s$ of the corresponding tangential components of electric and magnetic fields above and below the surface.

Radiation from the circulating surface wave is accompanied by a backward tilt of the equiphase surfaces, resulting in a departure from pure radial distribution of the field. The wave impedance of the "incident" field, looking radially out from the surface, transforms to the corresponding wave impedance at $r=\infty$ and consequently with a perfectly smooth transition in a homogeneous medium no "reflected" field would arise. In a practical case, however, there is likely to be some discontinuity outside the surface and although in those circumstances the "reflected" field itself may be of no great significance, its presence can have important consequences in increasing the power radiated. If we think of the wave pattern between adjacent equi-phase surfaces as stationary in space, the conditions closely resemble those of the E_{01} waveguide beyond cutoff, in which any "reflected" field from the termination provides for a flow of energy towards it. To investigate this effect in the present application, we will suppose that the surface wave field transforms to a plane wave

cylindrical wave and η may be positive or negative according to whether the power flows toward or away from the surface.

Also for simplicity, we shall suppose that the medium above the surface is air and that, compared with the real part, the imaginary part of ν is sufficiently small, to enable the latter to be neglected without causing serious error in writing down the equations to the field distribution. Before proceeding on this basis to a calculation of the power in the field, it is perhaps of interest to consider (5) in terms of the complex values of T and ν . Suppose that

$$j\nu = \alpha r + j\beta r; \quad (11)$$

then, it is clear that both α and β are inversely proportional to r when ν is constant.

From (2) and (11) we get

$$\left(T^2 - \frac{\nu^2}{r^2}\right) = \omega^2\mu\epsilon - j\omega\mu\sigma - (\beta^2 - \alpha^2) + j2\alpha\beta, \quad (12)$$

and, for the particular condition

$$\omega\mu\sigma = 2\alpha\beta, \quad (13)$$

the expression in (12) is purely real.

Unfortunately, this condition can only be established at one particular radius and consequently does not yield any simplified solution to the wave equation.

B. Calculation for the Power Radiated

The radial power density p_r is given by

$$p_r = \text{Re} [E_\theta H_y^*], \quad (14)$$

where H_y^* is the complex conjugate of H_y .

$$\rho = |\rho| e^{i\phi} = -\frac{H^-}{H^+} = -\frac{C_2 H_{-\nu}^{(1)}(Tr)}{C_1 H_{\nu}^{(2)}(Tr)} \quad (15)$$

and suppose that C_1 is real; then, using (7) and (8) with $\sigma=0$ and remembering that

$$J_{\nu}^2(Tr) + Y_{\nu}^2(Tr) = J_{-\nu}^2(Tr) + Y_{-\nu}^2(Tr) \quad (16)$$

with

$$Y_{\nu}(Tr)J_{\nu+1}(Tr) - J_{\nu}(Tr)Y_{\nu+1}(Tr) = \frac{2}{\pi Tr}, \quad (17)$$

we get from (14) at the surface where $r=r_s$,

$$p_r = \frac{2C_1^2}{\pi\omega\epsilon_0 r_s} \left[1 - |\rho_s|^2 + \left\{ \frac{\pi |\rho_s| Tr_s \sin \phi_s}{2} \right\} \{U + Q\} \right], \quad (18)$$

where

$$U = U_1 + U_2 = J_{\nu}(Tr_s)J_{\nu+1}(Tr_s) + Y_{\nu}(Tr_s) \cdot Y_{\nu+1}(Tr_s), \quad (19)$$

$$Q = Q_1 + Q_2 = J_{-\nu}(Tr_s)J_{-\nu-1}(Tr_s) + Y_{-\nu}(Tr_s)Y_{-\nu-1}(Tr_s), \quad (20)$$

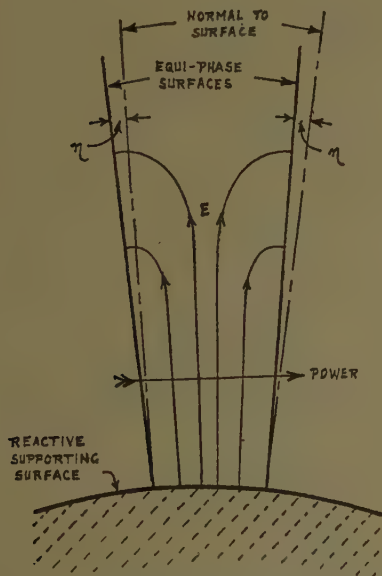


Fig. 2—Backward tilt of equiphase surfaces accompanying radiation.

at a finite height above the surface and that the reflection coefficient ρ_s at the supporting surface is proportioned to the angle of tilt $\pm\eta$ of the equi-phase surfaces, measured from the radial direction. (See Fig. 2.)

Thus for small angles,

$$\rho_s = -\rho_0\eta, \quad (10)$$

where ρ_0 is the reflection coefficient arising from the complete transformation of the surface wave to a plane

$$T = \omega \sqrt{\mu_0 \epsilon_0} = \frac{2\pi}{\lambda_0}, \quad (21)$$

and ρ_s is given by (10).

If, when looking radially outwards from the surface, the wave impedance in the absence of any reflected field is Z^+ and the free space wave impedance is Z_0 , we can write

$$\rho_0 = |\rho_0| e^{j\phi_0} = \frac{Z_0 - Z^+}{Z_0 + Z^+}. \quad (22)$$

Now

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} \quad (23)$$

and

$$Z^+ = R^+ + jX^+ = \frac{E_{\theta}^+}{H_y^+} = \frac{j}{\omega \epsilon_0 r_s} \left[\nu - Tr_s \frac{H_{\nu+1}^{(2)}(Tr_s)}{H_{\nu}^{(2)}(Tr_s)} \right], \quad (24)$$

where

$$R^+ = \frac{2}{\pi \omega \epsilon_0 r_s \{J_{\nu}^2(Tr_s) + Y_{\nu}^2(Tr_s)\}} \quad (25)$$

and

$$X^+ = - \left[\frac{T}{\omega \epsilon_0 \{J_{\nu}^2(Tr_s) + Y_{\nu}^2(Tr_s)\}} \right] \left[U - \frac{\nu}{Tr_s} \{J_{\nu}^2(Tr_s) + Y_{\nu}^2(Tr_s)\} \right]. \quad (26)$$

Since the surface impedance must have a very large inductive component (of value $-X^+$), we can neglect R^+ compared with X^+ and we then find

$$|\rho_0| = 1. \quad (27)$$

Within the range of interest near the surface the Liouville approximations to the Bessel functions can be applied because the condition

$$\frac{(Tr)^2[(Tr)^2 + 4\nu^2]}{4[\nu^2 - (Tr)^2]^3} \ll 1 \quad (28)$$

is easily satisfied in the cases of interest. Thus,

$$J_{\nu}(Tr_s) = \frac{1}{\nu!} \frac{e^{\nu(M-1)}}{\sqrt{M}} \left(\frac{Tr_s}{1+M} \right)^{\nu} \quad (29)$$

and

$$Y_{\nu}(Tr_s) = - \frac{\nu!}{\pi \nu} \frac{e^{\nu(1-M)}}{\sqrt{M}} \left(\frac{1+M}{Tr_s} \right)^{\nu}, \quad (30)$$

where

$$M = \sqrt{1 - (Tr_s/\nu)^2}; \quad (31)$$

and, for the large values of ν with which we are concerned,

$$V! = \left(\frac{\nu}{\pi} \right)^{\nu} \sqrt{2\pi\nu}. \quad (32)$$

If, for example, $\lambda_0 = 0.0314$ m. with $r_s = 1$ m, we have $T = 200$, and taking ν as one or two per cent greater than (Tr_s) , it is easy to see that $-Y_{\nu}(Tr_s) \gg J_{\nu}(Tr_s)$ so that

$$X^+ = - \frac{\nu}{\omega \epsilon_0 r_s} \left[\frac{e(1+M)}{e^M} - 1 \right], \quad (33)$$

or since $M \ll 1$,

$$X^+ \approx -1.72 \left(\frac{\nu}{\omega \epsilon_0 r_s} \right) = -1.72 \left(\frac{\beta}{\omega \epsilon_0} \right) \quad (34)$$

with

$$\phi_0 \approx /120^\circ \quad (35)$$

Taking η as a purely real angle,

$$|\rho_s| = \eta |\rho_0| \quad (10a)$$

and

$$\phi_s = \phi_0. \quad (36)$$

Using (27), (35), (10a), and (36) we can therefore re-write (18) as

$$p_r = \frac{2C_1^2}{\pi \omega \epsilon_0 r_s} \left[1 - \eta^2 - \left\{ \frac{\sqrt{3}\pi Tr_s \eta}{4} \right\} \{U + Q\} \right]. \quad (37)$$

The first term in this expression represents the power radiated as a result of the curvature of the surface, irrespective of any "reflected" field. The second term is the power thrown back by direct reflection of the surface wave field in its conversion to a free-space cylindrical wave, and will generally be quite negligible. The third term, which, as a rule, is also quite small but is sometimes important, provides for the additional radial power arising from the superposition of the "reflected" field on the "incident" field, as in a cutoff waveguide with a mismatched termination. It will be observed that since η is negative for a backward tilt of the equiphas surfaces, this condition corresponds to enhanced radiation as would be expected.

From (7) and (15) we have

$$H_y = [C_1 e^{-j\nu\theta} H_{\nu}^{(2)}(Tr)] [1 - |\rho| e^{j\phi}]. \quad (38)$$

Taking into consideration that ρ is proportional to η which is a very small angle, we can neglect the effect of $\rho = |\rho| e^{j\phi}$ in (38) on the tilt of the equiphas surfaces, and for the purpose of calculating η we can write

$$H_y = C_1 \sqrt{J_{\nu}^2(Tr) + Y_{\nu}^2(Tr)} \cdot e^{-j\nu(\theta+\tau)}, \quad (39)$$

where

$$\tan(\nu\tau) = \frac{+Y_{\nu}(Tr)}{J_{\nu}(Tr)} \quad (40)$$

and τ is the angle of deviation from pure radial distribution subtended at the center of curvature of the supporting surface.

Thus

$$\eta = r \frac{\partial \tau}{\partial r} = \frac{\partial}{\partial \nu} \left\{ \frac{\cos^2(\nu \tau)}{J_\nu^2(T r_s)} \right\}. \quad (41)$$

In order to get a better idea of the relative magnitudes of the terms in (37), it will be helpful to take a numerical example and compute the various quantities with the help of the Liouville approximations to the Bessel functions.

Thus, using (29), (30), (31), and (32), together with the corresponding expressions for the order $(\nu+1)$ and their counterparts of negative order, after inserting the appropriate values in (19) and (20), we find that

$$U_1 \approx \frac{1}{2\pi\nu M e} \left[\left\{ \frac{T r_s}{2\nu} \right\}^2 \left\{ 3 - \left(\frac{T r_s}{\nu} \right)^2 \right\} \cdot \left\{ 3 - \left(\frac{T r_s}{\nu+1} \right)^2 \right\} \right]^\nu, \quad (42)$$

$$U_2 \approx \frac{e}{\pi\nu M} \left[\left\{ \frac{\nu}{2T r_s} \right\}^2 \left\{ 1 + \left(\frac{T r_s}{\nu} \right)^2 \right\} \cdot \left\{ 1 + \left(\frac{T r_s}{\nu+1} \right)^2 \right\} \right]^\nu, \quad (43)$$

$$Q_1 \approx \frac{\nu U_2}{4T r_s} \left\{ 1 + \left(\frac{T r_s}{\nu+1} \right)^2 \right\} \{-1\}^{2\nu}, \quad (44)$$

$$Q_2 \approx \frac{2T r_s U_1}{(\nu+1)} \left\{ 3 - \left(\frac{T r_s}{\nu+1} \right)^2 \right\} \{-1\}^{2\nu}, \quad (45)$$

where use has been made of the approximations,

$$\frac{1+M}{e^M} \approx 1 - \frac{M^2}{2} \quad (46)$$

and

$$\frac{e^M}{1+M} \approx 1 + \frac{M^2}{2}, \quad (47)$$

where M has the value

$$M_1 = \sqrt{1 - \left(\frac{T r_s}{\nu+1} \right)^2}$$

for the Bessel functions of order $(\nu+1)$.

These approximations are only valid for $M \ll 1$, and over a comparatively small range of values of $(T r_s)$ relative to ν .

Suppose that we again consider the case of $\lambda_0 = 0.0314$ m (i.e., $f = 9.56 \times 10^9$ c/s) so that $T = 2\pi/\lambda_0 = 200$, and that we make calculations for $\nu = 205$, with $r_s = 1$ m and $r_s = 1.01$ m. Although it is difficult to get accurate numerical values, it is hoped that the calculations will be sufficient to show the relative importance of the different terms and the principal factors on which the radiated power depends.

$\nu = 205$:

a) When $r_s = 1$ m, (41) gives

$$\eta = -3.5 \text{ degrees at the surface}$$

and from (37) we find that

$$p_r = 1.2C_1^2[1 - 3.7 \times 10^{-3} + 0.55] = 1.86C_1^2.$$

b) When $r_s = 1.01$ m, then $\eta = -2.16$ degrees at the surface

and

$$p_r = 1.19C_1^2[1 - 1.4 \times 10^{-3} + 0.378] = 1.64C_1^2.$$

It will be observed that at the supporting surface, the angle of tilt η of the equiphase surfaces decreases as r_s increases and that the power radiated behaves in much the same way. Under the conditions postulated, the third term in (37) representing the component of the radial power density arising from the field reflected from outer space, is quite significant, but some of the quantities in this expression change so rapidly with r_s that a closer analysis is necessary to establish precisely the importance of the "reflected" field in this respect over a wider range of values.

C. Calculation for the Power Circulating around the Cylindrical Surface

The circumferential power outside the surface is given by

$$P_\theta = \int_{r=r_s}^{r=\infty} \text{Re} [-E_r H_\theta^*] dr \text{ per unit length in the } y \text{ direction}, \quad (48)$$

and, for the purpose of calculating this quantity, no great inaccuracy arises when we neglect the reflection terms in (7) and (9) for the field components.

Assuming that the medium outside the surface is air, we find that

$$P_\theta = \frac{\nu C_1^2}{\omega \epsilon_0} \int_{r=r_s}^{r=\infty} \left[\frac{1}{r} J_\nu^2(T r) + \frac{1}{r} Y_\nu^2(T r) \right] dr \quad (49)$$

and, therefore [5],

$$P_\theta = \frac{C_1^2}{2\omega \epsilon_0} \left[T r \left\{ J_{\nu+1}(T r) \frac{\partial}{\partial \nu} J_\nu(T r) - J_\nu(T r) \frac{\partial}{\partial \nu} J_{\nu+1}(T r) \right\} + J_\nu^2(T r) + T r \left\{ Y_{\nu+1}(T r) \frac{\partial}{\partial \nu} Y_\nu(T r) - Y_\nu(T r) \frac{\partial}{\partial \nu} Y_{\nu+1}(T r) \right\} + Y_\nu^2(T r) \right]. \quad (50)$$

In applying (50), we use the large argument approximations for the Bessel functions when $r \rightarrow \infty$ and the Liouville approximations together with those given by (46) and (47) when $r = r_s$. This leads to a rather long and involved expression for P_θ but it can be simplified without serious error if we remember that $\nu \approx T r_s$ to give

$$P_\theta = \frac{C_1^2}{\omega \epsilon_0} [1 - K], \quad (51)$$

where

$$K = \left[\frac{1}{2\pi\nu \sqrt{1 - \left(\frac{Tr_s}{\nu}\right)^2}} \right] \left[\left[\frac{3Tr_s}{2\nu} - \frac{1}{2} \left(\frac{Tr_s}{\nu}\right)^3 \right]^{2\nu} \cdot \left[\frac{1}{2} + \frac{Tr_s}{2e} \left\{ \frac{3 - \left(\frac{Tr_s}{\nu+1}\right)^2}{3 - \left(\frac{Tr_s}{\nu}\right)^2} \right\} \right] \cdot \left[\log_e \frac{3 - \left(\frac{Tr_s}{\nu}\right)^2}{3 - \left(\frac{Tr_s}{\nu+1}\right)^2} - \frac{1 - \frac{3}{4} \left(\frac{Tr_s}{\nu}\right)^2}{(\nu+1)} \right] \right] + \left[\frac{\nu}{2Tr_s} + \frac{Tr_s}{2\nu} \right]^{2\nu} \left[2 - e\nu \left\{ \frac{1 + \left(\frac{Tr_s}{\nu+1}\right)^2}{1 + \left(\frac{Tr_s}{\nu}\right)^2} \right\} \right] \cdot \log_e \left\{ \frac{\nu}{2Tr_s} + \frac{\nu Tr_s}{2(\nu+1)^2} \right\} \right] \quad (52)$$

For the case already discussed in which $\lambda_0 = 0.0314$ m, $T = 200$, $r_s = 1$ m, and $\nu = 205$, we find that $K = 0.0027$ and $P_\theta = 1.88 C_1^2$.

D. Rate of Attenuation of Circulating Wave Due to Radiation

Under the particular conditions considered, it is clear that reasonable approximations are obtained if, in place of (37) and (51), we write

$$\dot{p}_r \approx \frac{2C_1^2}{\pi\omega\epsilon_0 r_s} \quad (37a)$$

and

$$P_\theta \approx \frac{C_1^2}{\omega\epsilon_0} \quad (51a)$$

On this basis, the circumferential attenuation of the wave which arises from radiation (see Fig. 3) is

$$\alpha = \frac{1}{2} \log_e \left(\frac{P_\theta}{P_\theta - \dot{p}_r} \right); \quad (53)$$

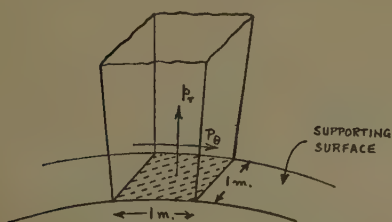


Fig. 3—Radial and circumferential components of power.

and, since

$$\frac{\dot{p}_r}{P_\theta} \approx \frac{2}{\pi r_s}, \quad (54)$$

we get for $r_s = 1$ m

$$\alpha = 4.4 \text{ db/m.}$$

Although (54) only applies over a very limited range of values of r_s , it is apparent that α increases rapidly as r_s decreases.

III. DISCUSSION ON THE EFFECT OF THE FIELD REFLECTED FROM OUTSIDE THE SURFACE IN RELATION TO THE POWER RADIATED

It is well known [1] that two highly reactive flat parallel surfaces separated by a distance which is large compared with the wavelength, will support waves of the Zenneck type in association with each surface (see Fig. 4). If the surfaces are both lossless, the power carried by the fields is entirely tangential, but when they have finite losses there will be a component of power directed toward each surface and the adjacent equiphase planes will be tilted forward. In addition, energy will flow towards the surfaces as a result of the reflected evanescent field which is incident on them. With only one surface, the Zenneck field would be expected to extend to an infinite distance from it and the presence of the second surface must clearly disturb that distribution of field in such a way as to set up a supplementary flow of energy which is in the nature of radiation from the first surface.

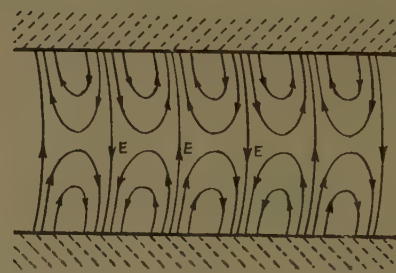


Fig. 4—Hybrid Zenneck-TEM wave between highly reactive parallel plane surfaces.

Between adjacent equiphase planes stretching across from one surface to the other, the field approximates an evanescent E mode of the kind supported by parallel metal plates and with a mismatched termination to produce a reflected field accompanied by the flow of energy towards it. Indeed, we can think of this mechanism as the means whereby radiation is set up by obstacles which, remote from the supporting surface, must necessarily disturb the smooth progress of the wave.

When the supporting surface becomes slightly cylindrical, the field pattern outside it remains of similar form but assumes a divergent distribution with the equiphase

surfaces tilted backwards by a small angle from the radial direction. As shown in Section II, the field is no longer purely evanescent and the conditions require a flow of energy outwards from the surface, irrespective of the existence of any reflected field.

IV. DISCUSSION ON THE EFFECT OF LOSSES IN THE SUPPORTING SURFACE

In order to support a surface wave on a cylindrical surface we know that it must be highly reactive so that a substantial proportion of the energy associated with the wave is stored within the surface. As the wave progresses, some of this energy passes across the supporting surface and is thereby radiated. There is, however, a considerable amount of energy also in the field outside the surface. When the medium below the surface is lossy, it must absorb power from the wave, and in the case of a flat surface, the process involves a forward tilt of the equiphase planes and a flow of energy across the interface into the surface. A very slight curvature of the surface tends to reverse that flow of energy and it seems reasonable to conclude that there must be a condition for which no energy passes in either direction across the supporting surface. Since, in these circumstances, there is bound to be some radiation as well as absorption of energy by the medium below the surface, there will be a progressive attenuation of the field both outside and inside the surface.

V. CONCLUSIONS

The problem of calculating the power radiated by a surface wave circulating around a cylindrical surface and the corresponding circumferential power is a difficult one because it leads to expressions that cannot eas-

ily be evaluated. By using some approximations, solutions have been obtained which should be applicable to a limited range of values. The analysis helps to elucidate the physical mechanism that characterizes the behavior of this form of wave.

VI. ACKNOWLEDGMENT

The author is indebted to his colleague, Dr. John Brown, for many helpful discussions on this subject.

REFERENCES

- [1] H. M. Barlow and A. L. Cullen, "Surface waves," *Proc. IEE*, vol. 100, pt. III, 329-347; November, 1953.
- [2] J. R. Wait, "Propagation of radio waves over a stratified ground," *Geophysics*, vol. 18, p. 416; April, 1953.
- S. A. Atwood, "Surface wave propagation over a coated plane conductor," *J. Appl. Physics*, vol. 22, p. 504; 1951.
- J. R. Wait, "Excitation of surface waves on conducting, stratified, dielectric-clad and corrugated surfaces," *J. Res. Natl. Bur. Standards*, RP 2807, vol. 59, p. 365; 1957.
- J. R. Wait, "Transmission and reflection of electromagnetic waves in the presence of stratified media," *J. Res. Natl. Bur. Standards*, vol. 61, pp. 205-232; September, 1958.
- A. L. Cullen, "The excitation of plane surface waves," *Proc. IEE*, vol. 101, pt. IV, pp. 225-234; February, 1954.
- J. R. Wait, "Radiation from a vertical dipole over a stratified ground," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-1, pp. 9-11, July, 1953; Part II, vol. AP-2, pp. 144-146, October, 1954.
- W. M. G. Fernando and H. M. Barlow, "An investigation of the properties of radial cylindrical surface waves launched over flat reactive surfaces," *Proc. IEE*, vol. 103B, pp. 307-318; May, 1956.
- B. Friedman and W. E. Williams, "Excitation of Surface Waves," *Proc. IEE*, Monograph No. 277 R, pp. 1-7; January, 1958.
- J. R. Wait and A. M. Conda, "Radiation from a slot on a large corrugated cylinder," *Proc. Symposium on the Propagation of Radio Waves, Liege, Belgium*, Academic Press, London, England; October, 1958.
- [3] R. S. Elliott, "Azimuthal surface waves on circular cylinders," *J. Appl. Phys.*, vol. 26, p. 368; 1955.
- [4] H. M. Barlow, "The power radiated by a surface wave circulating around a cylindrical surface," *Proc. IEE*, vol. 106B, pp. 180-185; March, 1959.
- [5] G. N. Watson, "Theory of Bessel Functions," Cambridge University Press, London, England; 1944.

Guiding of Electromagnetic Waves by Uniformly Rough Surfaces*

Part I

JAMES R. WAIT†

Summary—A simple derivation is given for the reflection of electromagnetic waves from a perfectly-conducting plane surface which has a uniform distribution of hemispherical bosses whose electrical constants are arbitrary. The spacing between the centers of the bosses is taken to be small, which is the justification for neglecting the incoherent radiation. An approximate boundary condition is developed which must be satisfied in an average sense by the tangential fields on the reference plane.

The excitation of surface waves on the rough surface is then discussed. It is indicated that to a first order, a rough surface of the kind described here possesses an inductive surface reactance and will support a trapped wave. The effect of finite conductivity of the bosses is to damp exponentially this trapped wave.

INTRODUCTION

MANY theoretical aspects of the reflection of electromagnetic and acoustic waves from rough, uneven, and corrugated surfaces have been discussed [1]–[10]. A favorite approach is to consider the surface as composed of a large number of independent elementary mirrors [11]. Models consisting of semicylindrical or hemispherical bosses on an infinite plane have been treated in extensive fashion by Twersky [12]–[15]. He also treats a model consisting of a random distribution of arbitrary protuberances on an infinite plane and takes into account multiple coherent scattering to obtain reflection coefficients [16]. By utilizing an energy balance relation, he then obtains differential cross sections on the assumption that the multiple incoherent scatter can be neglected. When the dimensions and the separations of the individual irregularities of the surface are small compared to the free space wavelength, and at near grazing incidence, the specular or coherent component of the scattered field is much larger than the diffuse or incoherent component. This limiting case has also been investigated by Biot [17], [18], who considered reflection of a plane wave from a flat perfectly conducting surface with a uniform and dense distribution of small hemispherical bosses. It is the prime purpose of the present paper to consider the response of such a surface to a dipole field, and to study the characteristics of the trapped waves which may exist near this surface.

THE SINGLE HEMISPHERICAL BOSS

It was suggested by Lord Rayleigh [1] that the field scattered by a semicylindrical boss on a perfectly conducting plane is the same as that scattered by a complete cylinder from both the incident wave and its geo-

metrical image with respect to the plane. This idea was extended to the hemispherical boss by Twersky [12]. In the present case, the boss is also considered to be hemispherical in form and is small compared to the wavelength. The incident field is locally uniform and induces an electric moment M^* in a direction parallel to the plane (see Fig. 1). More specifically,

$$M = Ea^3\Gamma \quad (1)$$

and

$$M^* = -\frac{H}{2}a^3\Gamma^* \quad (2)$$

where E and H are the resultant values of the incident and image fields at the center of the boss of radius a .

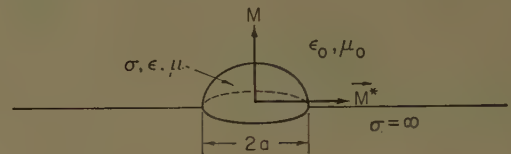


Fig. 1—Single hemispherical boss showing induced electric and magnetic dipoles.

Γ and Γ^* are factors which account for the electrical properties of the boss; they are chosen so that $\Gamma = \Gamma^* = 1$ for infinite conductivity. It is easily deduced from a previous analysis [19] or elsewhere [12], that

$$\Gamma = \frac{\left(\frac{\sigma + i\epsilon\omega}{i\epsilon_0\omega}\right) + \frac{A(g)}{2}}{\left(\frac{\sigma + i\epsilon\omega}{i\epsilon_0\omega}\right) - A(g)} \quad (3)$$

and

$$-\frac{\Gamma^*}{2} = \frac{\left(\frac{\mu}{\mu_0}\right) + \frac{A(g)}{2}}{\left(\frac{\mu}{\mu_0}\right) - A(g)}, \quad (4)$$

where

$$\begin{aligned} A(g) &= \frac{\sinh g - g \cosh g + g^2 \sinh g}{\sinh g - g \cosh g} \\ &= 1 + \frac{g^2}{1 - g \coth g}, \\ g &= [i\mu\omega(\sigma + i\epsilon\omega)]^{1/2}a; \end{aligned} \quad (5)$$

* The research reported in this paper has been sponsored largely by the AF Cambridge Research Center, Air Res. and Dev. Command, Bedford, Mass.

† National Bureau of Standards, Boulder, Colo.

σ , ϵ , and μ are the conductivity, dielectric constant and permeability of the homogeneous spherical boss, and ϵ_0 and μ_0 are the dielectric constant and the permeability of the surrounding lossless medium (*i.e.*, the air).

The above results are valid for $a \ll$ free space wavelength. If also $g \ll 1$, it follows that $A(g) \cong -2$; thus

$$\Gamma = \frac{(\epsilon - i\sigma/\omega) - \epsilon_0}{(\epsilon - i\sigma/\omega) + 2\epsilon_0} \quad (6)$$

and

$$-\frac{\Gamma^*}{2} = \frac{\mu - \mu_0}{\mu + 2\mu_0}, \quad (7)$$

which could be obtained directly from potential theory.

In most cases of interest, the permeability contrast is small (*i.e.*, $\mu \cong \mu_0$) and the conductivity of the boss material is sufficiently large (*i.e.*, $\sigma \gg \epsilon_0\omega$); thus (3) and (4) reduce to

$$\Gamma \cong 1 \quad (8)$$

and

$$\Gamma^* \cong 1 + \frac{3}{g^2} - \frac{3 \coth g}{g}, \quad (9)$$

where

$$g = (1 + i) \frac{a}{\delta}$$

and δ is the "skin-depth" defined by

$$\delta = \left(\frac{2}{\sigma\mu_0\omega} \right)^{1/2}.$$

It is seen that if $a \gg \delta$ (*i.e.*, $|g| \gg 1$),

$$\Gamma^* \cong 1 - \frac{3}{g} \cong 1. \quad (10)$$

If, on the other hand, $a \ll \delta$ (*i.e.*, $|g| \ll 1$), then

$$\Gamma^* \cong \frac{2g^2}{15} - \frac{2g^3}{315} \pm \text{terms in } g^4, g^5, \text{ etc.} \quad (11)$$

It can thus be concluded that for electrically small bosses of finite conductivity, the factor Γ corresponding to the strength of the induced electric dipole can be replaced by unity in most cases. However, the factor Γ^* corresponding to the strength of the induced magnetic dipole, in general, has a magnitude less than unity and is complex. A detailed numerical study of the induced magnetic dipole case has been given in a previous paper and the reader is referred to it for further information.¹

FORMULATION OF THE BOUNDARY CONDITIONS

It is now assumed that there is a uniform distribution of N hemispherical bosses per unit area (see Fig. 2). The

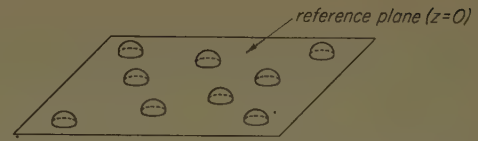


Fig. 2—A section of the surface containing a uniform distribution of hemispherical bosses.

electric and magnetic Hertz vectors Π and Π^* are given respectively by integrals over the surface of the reflecting plane.

$$\Pi = \iint NM(e^{-ikr}/r) da \quad (12)$$

and

$$\Pi^* = \iint NM^*(e^{-ikr}/r) da \quad (13)$$

where r is the distance from the observer to the variable surface element da . Denoting the scalar ϕ as any rectangular component of Π or Π^* , the integrals to contend with are of the form

$$\phi = \iint \rho(e^{-ikr}/r) da, \quad (14)$$

where ρ is the corresponding rectangular component of NM or NM^* . Following the suggestion of Biot, the distribution of ρ in a vanishingly small region near the plane behaves as a source of Newtonian potential. Thus, from potential theory or from integration of (14),

$$\left. \frac{\partial \phi}{\partial z} = -2\pi\rho \right]_{z=0} \quad (15)$$

where z is the coordinate normal to the surface. The corresponding vector forms are

$$\frac{\partial \Pi}{\partial z} = -2\pi NM \text{ and } \frac{\partial \Pi^*}{\partial z} = -2\pi NM^*. \quad (16)$$

For the distribution of bosses being considered here, (1) and (2) must be modified as follows:

$$M = (E + E')a^3\Gamma \quad (17)$$

and

$$M^* = -\frac{1}{2}(H + H')a^3\Gamma^* \quad (18)$$

where E' and H' represent the fields radiated by the dipoles themselves. The general boundary conditions at $z=0$ are then conveniently written

$$\frac{\partial \Pi}{\partial z} = -3v(E + E')\Gamma \quad (19)$$

and

$$\frac{\partial \Pi^*}{\partial z} = +\frac{3v}{2}(H + H')\Gamma^* \quad (20)$$

¹ The complex quantity Γ^* is to be identified with $\frac{2}{3}(\rho + iq)$ in the quoted reference [19], where numerical values of ρ and q are given as a function of μ/μ_0 and $(\sigma\mu_0\omega)^{1/2}a$.

where $v = 2\pi Na^3/3$ is the volume of hemispherical boss per unit area and has dimensions of length.

The boundary conditions stated by (19) and (20) were first given explicitly by Biot [17] although they are implied in the work of Twersky. [12] They are to be satisfied in an average sense over the reference plane which is conveniently taken as $z=0$. Furthermore, they are limited to small bosses since only the induced dipoles are considered. This is valid if the average dimensions of the bosses are small compared to the mutual spacing. For a dense distribution of bosses, the multipoles and the close range interaction may be important. Such effects have been considered by Biot [17] and will not be mentioned further in the present paper.

REFLECTION OF A VERTICALLY POLARIZED WAVE

A plane wave is considered to be incident at an angle θ on the plane $z=0$. The electric vector is contained in the plane of incidence; *i.e.*, the xz plane (see Fig. 3). The

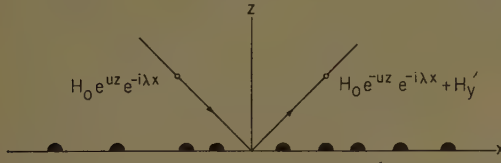


Fig. 3—The cross section in the xz plane showing the incident and total reflected H_y component.

field resulting from the incident and the reflected wave on a smooth perfectly conducting plane has only a y component of the magnetic field and is given by

$$H_y = H_0(e^{uz} + e^{-uz})e^{-i\lambda x} \quad (21)$$

where $u = ik \cos \theta$, $\lambda = k \sin \theta$, and H_0 is a constant. The corresponding component of the electric field is obtained from

$$E_z = \frac{1}{i\epsilon_0\omega} \frac{\partial H_y}{\partial x} \quad \text{and} \quad E_x = -\frac{1}{i\epsilon_0\omega} \frac{\partial H_y}{\partial z} \quad (22)$$

The field due the bosses can be derived from the Hertz vector Π , which has only a z component, and from Π^* , which has only a y component. Thus

$$H_y' = k^2 \Pi_y^* - i\epsilon_0\omega \frac{\partial \Pi_z}{\partial x} \quad (23)$$

The boundary equations (19) and (20) can now be written

$$\frac{\partial \Pi_z}{\partial z} = -\alpha(E_x + E_x') = \frac{-\alpha}{i\epsilon_0\omega} \left(\frac{\partial H_y}{\partial x} + \frac{\partial H_y'}{\partial x} \right) \quad (24)$$

and

$$\frac{\partial \Pi_y^*}{\partial z} = \frac{\alpha^*}{2} (H_y + H_y') \quad (25)$$

where $\alpha = 3v\Gamma$ and $\alpha^* = 3v\Gamma^*$. Differentiating both sides of (23) partially with respect to y and using (24) and (25) to eliminate Π_z and Π_y^* leads to

$$\frac{\partial}{\partial z} H_y' = \left(\alpha \frac{\partial^2}{\partial x^2} + \alpha^* \frac{k^2}{2} \right) (H_y + H_y'), \quad (26a)$$

which is the required boundary condition for the tangential magnetic field at $z=0$. Since $H_y + H_y'$ satisfies the wave equation, the boundary equation can be also written

$$\frac{\partial}{\partial z'} H_y' = \left(\alpha^* \frac{k^2}{2} - \alpha k^2 - \alpha \frac{\partial^2}{\partial z^2} \right) (H_y + H_y'). \quad (26b)$$

The coherent field scattered by the uniform distribution of bosses is also a plane wave. Therefore

$$H_y' = H_0' e^{-uz} e^{-i\lambda x}. \quad (27)$$

The boundary condition as stated by (26b) now leads readily to the relation

$$H_0 + H_0' = R_v H_0 \quad (28a)$$

where

$$R_v = \frac{2u - (\alpha^* k^2 - 2\alpha \lambda^2)}{2u + (\alpha^* k^2 - 2\alpha \lambda^2)} \quad (28b)$$

is a reflection coefficient. For an incident wave

$$H_0 e^{uz} e^{-i\lambda x} \quad (29)$$

the reflected (specular) wave is thus given by

$$H_0 R_v e^{-uz} e^{-i\lambda x}, \quad (30)$$

when it is remembered that $u = ik \cos \theta$ and $\lambda = k \sin \theta$. Numerical values of the amplitude and phase of R are readily computed from (28b). When the bosses are perfectly conducting,

$$\alpha^* = \alpha = 3v$$

and the resulting equation for R is agreement with Biot [17]. It is noted that at grazing incidence R approaches -1 , corresponding to a phase reversal of the reflected wave. In general, however, the finite conductivity of the bosses requires that the complex definition of $\alpha^* (= 3v\Gamma^*)$ be employed, although α can still be replaced by $3v$.

Eq. (28b) may also be obtained as a special case of Twersky's general result [16] when an appropriate form for the scattering function of a single boss is employed.

EXCITATION BY A DIPOLE

The previous solution for an incident vertically polarized wave can readily be generalized to the case of excitation by a vertical electric dipole. The behavior of the field for grazing angles and the existence of surface waves are then discussed.

With respect to a cylindrical coordinate system (ρ, ϕ, z) a vertical electric dipole, of length ds and average current I , is located at $z=h$ (see Fig. 4). The perfectly

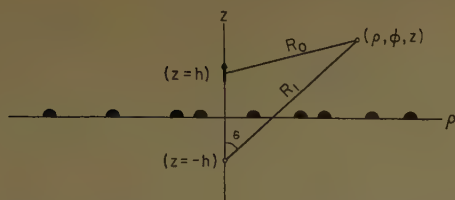


Fig. 4—A vertical electric dipole over the uniformly rough surface.

conducting plane surface, located at $z=0$, possesses a distribution of hemispherical bosses of radius a . The direct or primary magnetic field of the dipole and its image in the perfectly conducting plane has only a ϕ component H_ϕ , and is conveniently written

$$H_\phi = -b \frac{\partial}{\partial \rho} \psi$$

where

$$b = \frac{Ids}{4\pi} \quad \text{and} \quad \psi = \frac{e^{-ikR_0}}{R_0} + \frac{e^{-ikR_1}}{R_1} \quad (31)$$

with

$$R_0 = [(z-h)^2 + \rho^2]^{1/2} \quad \text{and} \quad R_1 = [(z+h)^2 + \rho^2]^{1/2}$$

This suggests that the field due to the bosses can be written in the form

$$H_\phi' = -b \frac{\partial}{\partial \rho} \psi' \quad (32)$$

where the scalar ψ' is a solution of the wave equation.

The function ψ can be written in the integral form [20]

$$\psi = \int_0^\infty [e^{-u|z-h|} + e^{-u(z+h)}] J_0(\lambda\rho) (\lambda/u) d\lambda \quad (33)$$

where $J_0(\lambda\rho)$ is a Bessel function of order zero with argument λ . As is well known, this integral can be interpreted as a spectrum of plane waves with angles of incidence θ which are generally complex, since $\lambda = k \sin \theta$ takes all values from 0 to ∞ . In further analogy to the previous section, the function ψ' is expressed in the form

$$\psi' = \int_0^\infty \psi_0'(\lambda) e^{-u(z+h)} J_0(\lambda\rho) (\lambda/u) d\lambda \quad (34)$$

where $\psi_0'(\lambda)$ is as yet an undetermined function of λ .

The boundary condition given by (26), developed for the tangential magnetic field, also holds for the functions ψ and ψ' :

$$\frac{\partial \psi'}{\partial z} = \left(\alpha^* \frac{k^2}{2} - \alpha k^2 - \alpha \frac{\partial^2}{\partial z^2} \right) (\psi + \psi') \quad \text{at } z=0. \quad (35)$$

Application of this boundary condition along with (33) and (34) leads readily to

$$1 + \psi_0'(\lambda) = R_v(\lambda) \quad (36)$$

where as expected, $R_v(\lambda)$ is given precisely by (28b), which was derived for plane wave incidence.

The total magnetic field H_ϕ^t is thus formally given by

$$H_\phi^t = -b \frac{\partial}{\partial \rho} \psi^t \quad (37)$$

where

$$\psi^t = \frac{e^{-ikR_0}}{R_0} + \int_0^\infty R_v(\lambda) e^{-u(z+h)} (\lambda/u) J_0(\lambda\rho) d\lambda. \quad (38)$$

The evaluation of the above integral is facilitated by converting $R_v(\lambda)$ in the following way (noting that $u^2 = \lambda^2 - k^2$):

$$R_v(\lambda) = \frac{2u - (\alpha^* k^2 - 2\alpha\lambda^2)}{2u + (\alpha^* k^2 - 2\alpha\lambda^2)} \quad (39a)$$

$$= - \frac{\left[u^2 + \frac{1}{\alpha} u + \left(1 - \frac{\alpha^*}{2\alpha} \right) k^2 \right]}{\left[u^2 - \frac{1}{\alpha} u + \left(1 - \frac{\alpha^*}{2\alpha} \right) k^2 \right]} \quad (39b)$$

$$= - \frac{[(u - ik\Delta_1)(u - ik\Delta_2)]}{[(u + ik\Delta_1)(u + ik\Delta_2)]} \quad (39c)$$

$$= -1 + \frac{2ik\Delta_1}{K(u + ik\Delta_1)} - \frac{2ik\Delta_2}{K(u + ik\Delta_2)} \quad (39d)$$

where

$$\Delta_1 = \frac{i}{2k\alpha} (1 + K), \quad \Delta_2 = \frac{i}{2k\alpha} (1 - K)$$

and

$$K = \left[1 - 4 \left(1 - \frac{\alpha^*}{2\alpha} \right) (k\alpha)^2 \right]^{1/2}.$$

This leads to

$$\psi^t = \frac{e^{-ikR_0}}{R_0} - \frac{e^{-ikR_1}}{R_1} + \psi_1 + \psi_2 \quad (40)$$

where

$$\psi_1 = \frac{2}{K} \int_0^\infty \frac{ik\Delta_1}{u(u + ik\Delta_1)} e^{-u(z+h)} \lambda J_0(\lambda\rho) d\lambda \quad (41)$$

$$\psi_2 = -\frac{2}{K} \int_0^\infty \frac{ik\Delta_2}{u(u + ik\Delta_2)} e^{-u(z+h)} \lambda J_0(\lambda\rho) d\lambda. \quad (42)$$

Now because of the initial assumption of small roughness, $|k\alpha| \ll 1$, and thus $|\Delta_1| \gg 1$. Eq. (41) for ψ_1 can now be evaluated by the conventional saddle point method to give

$$\psi_1 \cong \frac{2}{K} \left[\frac{\Delta_1}{\Delta_1 + C} + \frac{1}{2p_1(1 + C/\Delta_1)^3} + \frac{1.3}{(2p_1)^2(1 + C/\Delta_1)^5} + \dots \right] \times \frac{e^{-ikR_1}}{R_1} \quad (43)$$

subject to $kR_0 \gg 1$, and where

$$p_1 = \frac{-ik\Delta_1^2}{2} R_1 \quad \text{and} \quad C = \frac{z+h}{R_1}.$$

It should be noted that the contribution to the integral ψ_1 from the pole at $u = -ik\Delta_1$ or at $\lambda = \lambda_1 = -ik\sqrt{\Delta_1^2 - 1} \cong -ik\Delta_1$ leads to a term which varies as $\rho^{-1}e^{-ik\Delta_1(x+h)}e^{-k\Delta_1\rho}$, which is heavily damped in both the z and ρ directions and is thus of no physical interest. The terms containing inverse powers of p_1 can be neglected, since Δ_1^2 and kR are both large. Thus

$$\psi_1 \cong \frac{2}{K} \frac{\Delta_1}{\Delta_1 + C} \frac{e^{-ikR_1}}{R_1}. \quad (44)$$

The evaluation of the integral for ψ_2 is not quite so straightforward since, in view of the smallness of Δ_2 , the pole of the integrand at $u = -ik\Delta_2$ or at $\lambda = k\sqrt{1 - \Delta_2^2}$ is very near the branch point at $\lambda = k$. The necessary modifications of the saddle point method to treat general integrals of this type have been devised independently by Van der Waerden [21] and Clemmow [22] in 1950. In a previous paper by the author [20], the method was applied to an integral which is identical in form to ψ_2 . It thus follows that

$$\psi_2 = -\frac{2}{K} \frac{\Delta_2}{\Delta_2 + C} [1 - F(w)] \frac{e^{-ikR_1}}{R_1} \quad (45)$$

where

$$w = -\left(1 + \frac{C}{\Delta_2}\right)^2 \frac{ik\Delta_2^2 R_1}{2} \quad (46)$$

and

$$F(w) = 1 - i(\pi w)^{1/2} e^{-w} \operatorname{erfc}(iw^{1/2}) \quad (47)$$

where

$$\operatorname{erfc}(Z) = \frac{2}{\pi^{1/2}} \int_Z^\infty e^{-x^2} dx \quad (48)$$

is the complement of the error function of (complex) argument Z . The above formula for ψ_2 is valid at large distances (*i.e.*, $kR \gg 1$) and for reasonably small values of Δ_2 (*i.e.*, $|\Delta_2| < 0.1$ say). Furthermore, the real part of Δ_2 is to be positive although it can be vanishingly small.

The final result thus reads

$$\psi^t = \frac{e^{-ikR_0}}{R_0} + \left[\frac{2}{K} \frac{\Delta_1}{\Delta_1 + C} - 1 - \frac{2}{K} \frac{\Delta_2}{\Delta_2 + C} [1 - F(w)] \right] \frac{e^{-ikR_1}}{R_1}. \quad (49)$$

In the case of vanishing roughness $|\Delta_1| \rightarrow \infty$ and $|\Delta_2| \rightarrow 0$; thus $F(w) \rightarrow 1$ which leads to

$$\psi^t = \frac{e^{-ikR_0}}{R_0} + \frac{e^{-ikR_1}}{R_1} \quad (50)$$

as it should. Eq. (49) is also simplified somewhat in the case when the source dipole and the receiver are near the surface such that C can be replaced by zero. Then

$$\psi^t = \frac{e^{-ik\rho}}{\rho} \frac{2}{K} F(p) \quad \text{where} \quad p = -\frac{ik}{2} \Delta_2^2 \rho. \quad (51)$$

This equation is completely analogous to the case of propagation over a flat homogeneous earth. The analogy is complete if Δ_2 is replaced by $Z/120\pi$ where Z is the surface impedance on the flat ground plane [20]. This supports the idea that a uniformly rough surface of the type described can be characterized, at least in a partial sense, by a surface impedance. An alternate derivation of (51) is given in the Appendix.

DISCUSSION OF THE SURFACE WAVE

To illustrate the nature of the general solution and to demonstrate the existence of a surface wave an asymptotic solution is now developed. When $|w| \rightarrow \infty$,

$$F(w) \cong -2i(\pi w)^{1/2} e^{-w} - \frac{1}{2w} - \frac{3}{2^2 w^2} - \frac{3 \cdot 5}{2^3 w^3} - \frac{3 \cdot 5 \cdot 7}{2^4 w^4} \dots \quad (52)$$

Thus in the asymptotic sense

$$\begin{aligned} \psi^t \cong & \frac{e^{-ikR_0}}{R_0} + R_v \frac{e^{-ikR_1}}{R_1} \\ & - \frac{2}{K} \left(\frac{\Delta_2}{\Delta_2 + C} \right) \left[\frac{1}{2w} + \frac{3}{2^2 w^2} + \frac{3 \cdot 5}{2^3 w^3} \right. \\ & \quad \left. + \frac{3 \cdot 5 \cdot 7}{2^4 w^4} + \dots \right] \frac{e^{-ikR_1}}{R_1} \\ & - i \frac{4}{K} \left(\frac{\Delta_2}{\Delta_2 + C} \right) (\pi w)^{1/2} e^{-w} \frac{e^{-ikR_1}}{R_1} \end{aligned} \quad (53)$$

where

$$R_v = \frac{2}{K} \frac{\Delta_1}{\Delta_1 + C} - 1 - \frac{2}{K} \frac{\Delta_2}{\Delta_2 + C}. \quad (54)$$

R_v is the same specular coefficient as in (28b) since $C = \cos \theta$.

The third term in the expression for ψ^t is of the nature of a surface wave. For small values of C this term predominates if the losses on the surface are small. For example, if

$$\Delta_2 = i|\Delta_2| \quad (55)$$

corresponding to

$$\Gamma = \Gamma^* = 1, \text{ the third term behaves like} \quad (56)$$

$$\begin{aligned} & \rho^{-1/2} \exp \left[-ik \left(1 + \frac{|\Delta_2^2|}{2} \right) \rho \right] \\ & \cdot \exp [-k|\Delta_2|(h+z)]. \end{aligned} \quad (57)$$

This illustrates that the surface wave component has a phase velocity which is $(1 + |\Delta_2|^2/2)^{-1}$ times the free space velocity and the wave is heavily damped in the vertical direction.

To illustrate the influence of losses the defining equation for Δ_2 can be written

$$\Delta_2 = \frac{i}{6kv} \left[1 - \sqrt{1 - 4 \left(1 - \frac{\Gamma^*}{2} \right) (3kv)^2} \right] \quad (58a)$$

if the bosses are of finite conductivity and are non magnetic; i.e., $\alpha = 3v$ and $\mu = \mu_0$. Since kv is small, the equation can be further approximated by

$$\Delta_2 \cong i \left(1 - \frac{\Gamma^*}{2} \right) 3kv. \quad (58b)$$

For very high conductivity or for higher frequencies (i.e., $\delta \ll a$), the skin depth δ may be small compared to the dimensions of the bosses; then

$$\Gamma^* \cong 1 - \frac{3}{g} \cong 1 + \frac{3}{2} (i - 1) \frac{\delta}{a}$$

and consequently

$$\Delta_2 \cong i \frac{(3kv)}{2} \left[1 - i \frac{3}{2} \frac{\delta}{a} \right]. \quad (59)$$

The surface wave term now behaves like

$$\rho^{-1/2} \exp \left[-ik \left(1 + \frac{9(kv)^2}{8} \right) \rho \right] \cdot \exp \left[-\frac{27}{8} (kv)^2 k \rho \frac{\delta}{a} \right] \exp \left[-\frac{3k^2 v}{2} (h + z) \right], \quad (60)$$

which is exponentially damped in the ρ direction.

The other special case of interest is at the low frequencies where the dimensions of the conducting boss are small compared to the skin depth; then

$$\Gamma^* \cong + \frac{2i}{15} \left(\frac{a}{\delta} \right)^2 \quad (61)$$

so that

$$\Delta_2 \cong i3kv \left[1 - \frac{i}{15} \left(\frac{a}{\delta} \right)^2 + \text{higher terms in } \left(\frac{a}{\delta} \right) \right]. \quad (62)$$

The corresponding behavior of the surface wave is thus

$$\rho^{-1/2} \exp \left[-ik \left(1 + \frac{9}{2} (kv)^2 \right) \rho \right] \cdot \exp \left[-\frac{3}{5} (kv)^2 k \rho \left(\frac{a}{\delta} \right)^2 \right] \exp (-3k^2 v (h + z)), \quad (63)$$

which is also exponentially damped in the ρ direction.

When the radius a of the bosses is comparable to the skin depth, it is necessary to employ the more general

relation for Δ_2 which, for $kv \ll 1$, can be written in the form

$$\Delta_2 \cong kvq + i3kv \left(1 - \frac{p}{3} \right) \quad (64)$$

where p and q are tabulated as a function of $(\sigma\mu_0\omega)^{1/2}a$ (or $\sqrt{2}a/\delta$) and μ/μ_0 in a previous paper [19]. For the whole range of a/δ the quantities p and q are real and positive. Thus, in general, the effect of finite conductivity is to attenuate significantly the surface wave, whereas it will not appreciably modify the specular component.

EXTENSION TO HORIZONTAL POLARIZATION

The above analysis deals exclusively with vertical polarization. For plane-wave incidence, the extension to horizontal polarization is simple, as pointed out by Twersky [16] and Biot [17]. In the case of a plane wave incident on a perfectly conducting plane surface at $z = 0$, the electric field is given by

$$E_y = E_0 [e^{uz} - e^{-uz}] e^{-i\lambda x} \quad (65)$$

and $E_x = E_z = 0$. As before, $u = ik \cos \theta$ and $\lambda = k \sin \theta$ for an angle of incidence θ . The coherent wave which results from the uniform distribution of hemispherical bosses is written

$$E_y' = E_0' e^{-uz} e^{-i\lambda x} \quad (66)$$

where E_0' is yet to be determined. Only induced magnetic dipoles are now present, since the electric dipoles induced are cancelled by the conducting plane surface. Thus E_y' can be derived from a magnetic Hertz vector which has only an x component Π^* , in the manner

$$E_y' = -i\mu_0\omega \frac{\partial \Pi_x^*}{\partial z}. \quad (67)$$

The specific boundary condition as obtained from (20) now reads

$$\frac{\partial \Pi_x^*}{\partial z} = \frac{\alpha^*}{2} (H_x + H_x') \text{ at } z = 0. \quad (68)$$

Since

$$H_x = \frac{1}{i\mu_0\omega} \frac{\partial E_y}{\partial z}$$

and in view of (67) it readily follows that

$$E_y' = -\frac{\alpha^*}{2} \left(\frac{\partial E_y}{\partial z} + \frac{\partial E_y'}{\partial z} \right) \text{ at } z = 0 \quad (69)$$

is the required boundary condition. Application of (69) to (65) and (66) leads immediately to

$$E_0' - E_0 = R_h E_0 \quad (70)$$

where R_h is the reflection coefficient given explicitly by

$$R_h = \frac{u\alpha^* + 2}{u\alpha^* - 2} \quad (71)$$

where $\alpha^* = 3v\Gamma^*$. The results for normal incidence may be checked by noting that, for $\theta = 0$

$$R_h = R_v = \frac{ik\alpha^* + 2}{ik\alpha^* - 2}. \quad (72)$$

Since α^* is small, R_v is very near -1 for all angles.

The corresponding solution for a small horizontal loop or vertical magnetic dipole over the rough surface is readily carried out by using the same method as above for the electric dipole. For a small loop of area dA and carrying a current I , the resulting electric field has only a ϕ component E_ϕ^t , and is given formally by

$$E_\phi^t = -\frac{i\mu_0\omega IdA}{4\pi} \frac{\partial}{\partial \rho} \Phi^t \quad (73)$$

where

$$\Phi^t = \frac{e^{-ikR_0}}{R_0} + \int_0^\infty R_h(\lambda) e^{-u(z+h)} (\lambda/u) J_0(\lambda\rho) d\lambda, \quad (74)$$

where

$$R_0 = [(z-h)^2 + \rho^2]^{1/2},$$

and where $R_h(\lambda)$ is given by (72). Evaluation of this integral by the saddle point method leads readily to

$$\Phi^t \cong \frac{e^{-ikR_0}}{R_0} - \frac{2 + ik\alpha^*C}{2 - ik\alpha^*C} \frac{e^{-ikR_1}}{R_1} \quad (75)$$

where

$$R_1 = [(z-h)^2 + \rho^2]^{1/2} \text{ and } C = \frac{z+h}{R_1}.$$

In view of the smallness of α^* , this can be written in the form

$$\Phi^t \cong \frac{e^{-ikR_0}}{R_0} - (1 + \nu) \frac{e^{-ikR_1}}{R_1} \quad (76)$$

where ν is a small quantity given by

$$\nu \cong ik\alpha^*C = 3ikv\Gamma^*C. \quad (77)$$

For highly conducting bosses, $\Gamma^* \cong 1$; thus the main effect of the roughness is to advance the phase of the reflected wave by $3kvC$ radians.

It is noted that in the case of horizontal polarization there is no surface wave, and for oblique incidence the influence of roughness (within the scope of the assumed model) is generally much smaller than for vertical polarization.

CONCLUSION

The above analysis considers only the coherent part of the reflection from a rough surface. The neglect of the

incoherent or scattered radiation is always justified if the mutual spacing between the bosses is small compared to the wavelength [23]. In fact, for highly oblique incidence and surface waves, the secondary radiation from bosses separated by distances even comparable to the wavelength are also in coherence. Further support to this concept is given in the appendix where an alternative derivation is carried out.

The present analysis is being extended to curved convex surfaces which possess a roughness in the sense described above. The propagation of waves between concentric rough surfaces is also being considered. Formally, the method is to apply the boundary conditions stated by (26a) and (26b) as generalized to the appropriate orthogonal coordinate systems. An example of this approach is used in the treatment of the radiation from a magnetic line source in the presence of a corrugated cylinder [30].

APPENDIX

INTEGRAL EQUATION APPROACH

As a matter of interest, an alternative derivation is now carried out which yields an integral equation for the attenuation function for wave propagation over a rough surface. The source is considered to be a vertical electric dipole which is in close proximity to the $z=0$ plane. Over this plane is a distribution of electrically small hemispherical bosses of finite conductivity. The vertical field near the $z=0$ plane and averaged over this plane can be written in the form

$$E_z^t(x, y) = \frac{i\mu\omega Ids}{2\pi\rho} e^{-ik\rho} F(\rho) \quad (78)$$

where Ids is the current moment of the source dipole, $F(\rho)$ is an unknown attenuation function, and $\rho = (x^2 + y^2)^{1/2}$. For a perfectly conducting flat plane, in the absence of bosses, $F(\rho)$ would be unity.

The quantity $F(\rho)$ is assumed to vary slowly within a wavelength and also within a distance which includes many obstacles. The total field e_s^t scattered by one obstacle can be expressed in the form

$$e_s^t(x', y') = \frac{k^2 e^{-ikr'}}{r'} a^3 \left[1 - \frac{\Gamma^*}{2} \mathbf{i}_r' \cdot \mathbf{i}_\rho' \right] E_s^t(x', y') \quad (79)$$

where \mathbf{i}_r' and \mathbf{i}_ρ' are unit vectors in the r' and ρ' directions respectively (see Fig. 5). Eq. (79) is obtained by

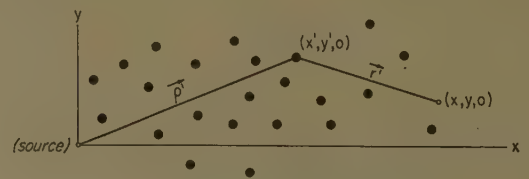


Fig. 5—The xy plane and the coordinate system for the integral equation approach.

considering the electric and magnetic dipole moments induced in the boss by the total average field. Γ^* is defined by (7), and as mentioned it can be replaced by unity for perfect conductivity. The field $E_x'(x, y)$ scattered by all the bosses is then

$$E_x'(x, y) = \iint_s N(x, y) e_s'(x', y') dx' dy' \quad (80)$$

where the integration is carried out over the whole ground plane and where $N(x', y')$ is the distribution of obstacles per unit area which can be regarded as a function of x' and y' . The total average field is thus²

$$E_x^t(x, 0) = \frac{i\mu\omega Ids}{2\pi\rho} e^{-ik\rho} + E_x'(x, 0) \quad (81)$$

which, on eliminating common factors, is transformed to

$$F(x, 0) = 1 + k^2 \iint N(x', y') \frac{e^{-ik(r'+\rho')}}{r'\rho'} \times a^3 \left[1 - \frac{\Gamma^*}{2} \mathbf{i}_{r'} \cdot \mathbf{i}_{\rho'} \right] F(x', y') dx' dy' \quad (82)$$

where

$$\rho' = [(x')^2 + (y')^2]^{1/2} \text{ and } r' = [(x - x')^2 + (y')^2]^{1/2}.$$

This is a two-dimensional integral equation for the attenuation function $F(x, 0)$. The major contribution to the double integral occurs when the phase of the exponential term is nearly constant, since the others are relatively slowly varying. With this in mind, the exponent is expanded in a power series in y^2 as follows:

$$r' + \rho' \cong x' + |x - x'| + \frac{(y')^2}{2} \left[\frac{1}{x'} + \frac{1}{|x - x'|} \right] \quad (83)$$

where terms containing y^4 , y^6 are neglected. If $0 < x' < x$, it is seen that

$$r' + \rho' \cong x + \frac{(y')^2}{2} \left[\frac{1}{x'} + \frac{1}{x - x'} \right], \quad (84)$$

but if $x' < 0$ or if $x' > x$, the quantity $r' + \rho'$ is rapidly varying with x' . The integration over x' can thus be taken from 0 to x . The integration over y' requires evaluating integrals of the form

$$I(x) = \int_{-\infty}^{+\infty} P(x', y') e^{-i\alpha^2(y')^2} dy' \quad (85)$$

where

$$\alpha^2 = \frac{k}{2} \left[\frac{1}{x'} + \frac{1}{x - x'} \right]$$

and where $P(x', y')$ is a slowly varying function of y' . It is clear that $I(x')$ can be approximated by

$$I(x') \cong P(x', 0) \int_{-\infty}^{+\infty} e^{-i\alpha^2(y')^2} dy' \cong \frac{P(x', 0)}{\sqrt{i\alpha}} \int_{-\infty}^{+\infty} e^{-Y^2} dY = \sqrt{\frac{\pi}{i}} \frac{P(x', 0)}{\alpha} \quad (86)$$

Using these results, the two-dimensional integral equation is now reduced to the one-dimensional integral equation

$$F(x) = 1 - \left(\frac{ikx}{2\pi} \right)^{1/2} \int_0^x \frac{\Delta(x') F(x - x')}{[x'(x - x')]^{1/2}} dx' \quad (87)$$

where $F(x) = F(x, 0)$ and where

$$\Delta(x') = i2\pi k a^3 N(x', 0) \left(1 - \frac{\Gamma^*}{2} \right)$$

is a function which is a measure of the roughness along a line connecting the source dipole and the observer.

When the distribution function $\Delta(x')$ is a constant it becomes identical in form to the quantity Δ_2 as defined by (58b). The resulting integral equation now has the form

$$F(x) = 1 - \left(\frac{ikx}{2\pi} \right)^{1/2} \Delta_2 \int_0^x \frac{F(x - x')}{[x'(x - x')]^{1/2}} dx'. \quad (88)$$

This integral equation has been studied extensively [24]–[29]. From the published work it immediately follows that

$$F(x) = 1 - i(\pi p)^{1/2} e^{-p} \operatorname{erfc}(ip^{1/2}) \quad (89)$$

where

$$p = -\frac{ik}{2} \Delta_2^2 x$$

which is equivalent to (47) and (51).

The integral equation approach thus gives results for the attenuation characteristics which are essentially identical to the early method, provided kv is small compared to unity. It is rather interesting to note that in the earlier derivation it was necessary to assume that the spacing between the bosses is small compared to wavelength. From the latter results, however, this restriction appears to be overly stringent for grazing or highly oblique incidence. The restriction on boss spacing for grazing incidence is that the attenuation function should vary slowly in a distance which includes many obstacles.

ACKNOWLEDGMENT

The author thanks H. Dougherty for his very helpful comments on Part I.

² There is no loss in generality in taking $y=0$.

BIBLIOGRAPHY

- [1] Lord Rayleigh, "On the light dispersed from fine lines ruled upon reflecting surfaces," *Philosophical Magazine*, vol. 14, pp. 350-359; September, 1907.
- [2] S. O. Rice, "Reflections of electromagnetic waves from slightly rough surfaces," *Commun. on Pure and Appl. Math.*, vol. 4, pp. 351-378; August, 1951.
- [3] J. W. Miles, "On nonspecular reflection at a rough surface," *J. Acoust. Soc. Am.*, vol. 26, pp. 191-199; March, 1954.
- [4] J. R. Wait, "Scattering of electromagnetic waves from a 'lossy' strip on a conducting plane," *Can. J. Phys.*, vol. 33, pp. 383-390; July, 1955, and "Radiation from an electric dipole in the presence of a corrugated cylinder," *Appl. Sci. Res.*, vol. B6, No. 1-2, pp. 117-123; 1956.
- [5] K. Norton, "Transmission loss of space waves propagated over irregular terrain," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-3, pp. 152-166; August, 1952.
- [6] Y. P. Lysanov, "On the scattering of electromagnetic waves from a rough surface," *Doklady Akad. Nauk (USSR)*, vol. 87, No. 5, pp. 719-722; 1952.
- [7] W. C. Hoffmann, "Scattering of e.m. waves from a rough surface," *Quart. Appl. Math.*, vol. 13, pp. 291-304; October, 1955.
- [8] L. M. Brekhovskikh, "The diffraction of waves by a rough surface, Parts I and II," *Zhurnal Eksperimental'noi Teoreticheskoi Fiziki, (USSR)*, vol. 23, No. 3(9), pp. 275-288, 1952, and vol. 23, pp. 289-304, 1952.
- [9] W. S. Ament, "Towards a theory of reflection by a rough surface," *PROC. IRE*, vol. 41, pp. 142-146; January, 1953, and "Forward and back-scattering from certain rough surfaces," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-4, pp. 369-373; July, 1956.
- [10] P. Beckmann, "A new approach to the problem of reflection from a rough surface," *Acta Technica CSAV*, vol. 2, No. 4, pp. 311-355; 1957.
- [11] H. Davies, "The reflection of e.m. waves from a rough surface, Part IV," *Proc. IEE*, vol. 101, pp. 209-214; August, 1954.
- [12] V. Twersky, "On the non-specular reflection of electromagnetic waves," *J. Appl. Phys.*, vol. 22, pp. 825-835; June, 1951.
- [13] V. Twersky, "Multiple scattering of radiation by an arbitrary planar configuration of parallel cylinders and by two parallel cylinders," *J. Appl. Phys.*, vol. 23, pp. 407-414; April, 1952.
- [14] V. Twersky, "Certain transmission and reflection theorems," *J. Appl. Phys.*, vol. 25, pp. 859-862; July, 1954.
- [15] V. Twersky, "Scattering theorems for bounded periodic structures," *J. Appl. Phys.*, vol. 27, pp. 1118-1122; October, 1956.
- [16] V. Twersky, "On scattering and reflection of electromagnetic waves by rough surfaces," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-5, pp. 81-89; January, 1957.
- [17] M. A. Biot, "Some new aspects of the reflection of e.m. waves on a rough surface," *J. Appl. Phys.*, vol. 28, pp. 1455-1463; December, 1957.
- [18] M. A. Biot, "On the reflection of e.m. waves on a rough surface," *J. Appl. Phys.*, vol. 29, p. 998; June, 1958.
- [19] J. R. Wait, "Complex magnetic permeability of spherical particles," *PROC. IRE*, vol. 41, pp. 1664-1667; November, 1953.
- [20] J. R. Wait, "Excitation of surface waves on conducting, stratified, dielectric-clad, and corrugated surfaces," *J. Res. Natl. Bur. Standards*, vol. 59, pp. 365-377; December, 1957.
- [21] B. L. Van der Waerden, "On the method of saddle points," *Appl. Sci. Res.*, vol. B2, No. 1, pp. 33-43; 1950.
- [22] P. C. Clemmow, "Some extensions to the method of integration by steepest descents," *Quart. J. Mech. Appl. Math.*, vol. 3, pt. 2, pp. 241-260; 1950.
- [23] M. P. Bachynski, "Microwave Propagation Over Rough Surfaces," Res. Labs., RCA Victor Co. Ltd., Montreal, Canada, Res. Rept. No. 7-150-1; June, 1958.
- [24] G. A. Grunberg and V. A. Fock, "On the Theory of Coastal Refraction of Electromagnetic Waves," edited by B. A. Vvedensky, Academy of Sciences U.S.S.R., Moscow, p. 69; 1948.
- [25] E. L. Feinberg, "Propagation of radio waves along an imperfect surface," *J. Phys. U.S.S.R.*, vol. 10, No. 5, pp. 410-418; 1946.
- [26] H. Bremmer, "The extension of Sommerfeld's formula for the propagation of radio waves," *Physica*, vol. 20, pp. 441-458; August, 1954.
- [27] G. C. Hufford, "An integral equation approach to the problem of wave propagation over an irregular surface," *Quart. Appl. Math.*, vol. 9, pp. 391-404; January, 1952.
- [28] J. R. Wait, "Mixed path ground wave propagation," *J. Res. Natl. Bur. Standards*, vol. 57, pp. 1-15; July, 1956, and vol. 59, pp. 19-26; July, 1957.
- [29] Z. Godzinski, "The Use of Secondary Sources in the Theory of Ground Wave Propagation Over an Inhomogeneous Earth," IEE Monograph 299R; April, 1958.
- [30] J. R. Wait and A. M. Conda, "Radiation from a slot on a large corrugated cylinder," *Proc. Symposium on Radio Wave Propagation*, Liège, Belgium, October, 1958. (To be published by Academic Press, London, England, 1959.)

Guiding of Electromagnetic Waves by Uniformly Rough Surfaces*

Part II

JAMES R. WAIT†

Summary—Using the model of a single rough surface described in Part I, the influence of curvature is considered. The starting point is a residue-series solution for a vertical dipole over a sphere with an arbitrary surface impedance. It is shown that the curvature has a profound influence on the nature of the surface wave, although it uniformly approaches the conditions for a plane boundary as the radius of curvature approaches infinity.

The propagation between the space bounded by two parallel uniformly rough surfaces is also considered. This is treated first for plane boundaries and then for concentric spherical boundaries. The latter model is useful in explaining certain experimental data on the terrestrial propagation of VLF radio waves.

INTRODUCTION

IN this sequel the influence of curvature on the propagation of waves along a uniformly rough surface is briefly discussed. The model chosen is a spherical surface which has a distribution of hemispherical bosses. The guiding of waves in a parallel plate region with uniformly rough walls also is considered. Finally the guiding of waves in a region bounded by two concentric uniformly rough surfaces is mentioned. Because of limitations of space the solutions are only outlined.

SURFACE WAVES ON A CURVED BOUNDARY WITH UNIFORM ROUGHNESS

The boundary condition stated by (26a) or (26b) in Part I may be applied to any surface whose mean curvature is small. For example, if the plane boundary considered in Part I is now imagined to be a spherical boundary with a distribution of hemispherical bosses, the relevant boundary condition reads

$$\frac{\partial}{\partial r} H_{\phi}^t = \left(\alpha^* \frac{k^2}{2} - \alpha k^2 - \alpha \frac{\partial^2}{\partial r^2} \right) H_{\phi}^t \Big|_{r=a} \quad (90)$$

where the spherical coordinate system is (r, θ, ϕ) and the spherical reference boundary is $r=a$ (see Fig. 6).

Using the method of separation of variables, the resultant field for a radial electric dipole at $r=b$ may be expressed in the following manner:

$$H_{\phi}^t = \frac{ikIds}{4\pi} \frac{\partial U}{\partial \theta} \quad (91)$$

where

$$U = U_s + U_e$$

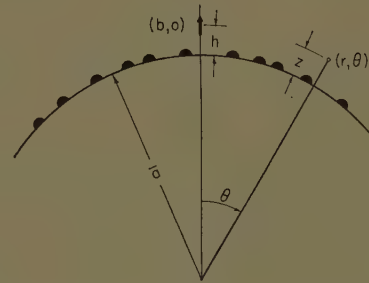


Fig. 6—Vertical dipole over a spherical surface which has a uniform distribution of bosses.

and

$$U_s = \sum_{n=0}^{\infty} (2n+1) P_n(\cos \theta) \begin{cases} h_n^{(1)}(kr) h_n^{(2)}(kb); & r < b \\ h_n^{(2)}(kr) h_n^{(1)}(kb); & r > b \end{cases} \quad (92)$$

has the proper source singularity;

$$U_s = \sum_{n=0}^{\infty} (2n+1) A_n h_n^{(2)}(kr) P_n(\cos \theta) \quad (93)$$

is the secondary field and satisfies the boundary condition at $r=a$ if

$$A_n \cong - \frac{h_n^{(1)}(ka)}{h_n^{(2)}(ka)} \frac{[\ln' x h_n^{(1)}(x) - i\Delta_1]}{[\ln' x h_n^{(2)}(x) - i\Delta_1]} \cdot \frac{[\ln' x h_n^{(1)}(x) - i\Delta_2]}{[\ln' x h_n^{(2)}(x) - i\Delta_2]}_{x=ka} \quad (94)$$

In the above \ln' is the logarithmic derivative, Δ_1 and Δ_2 are as defined in Part I, and $h_n^{(1)}(x)$ and $h_n^{(2)}(x)$ are spherical Hankel functions of the first and second kind respectively, of order n and argument x .

The Watson transformation [31] is now applied to these harmonic type series and this results in a residue series representation (without a remainder integral because $A_{n-1/2} = A_{-n-1/2}$). The function U is thus proportional to

$$\sum_{\nu} \frac{(\nu+1) h_{\nu}^{(2)}(kb) h_{\nu}^{(2)}(kr) P_{\nu}[\cos(\pi - \theta)]}{\sin \pi \nu M'(\nu) [h_{\nu}^{(2)}(ka)]^2} \quad (95)$$

where $M'(\nu) = [dM(y)/dy]_{y=\nu}$ and ν is a solution of $M(x) = 0$ where

$$M(\nu) = [\ln' x h_{\nu}^{(2)}(x) - i\Delta_1][\ln' x h_{\nu}^{(2)}(x) - i\Delta_2], \quad (96)$$

evaluated at $x=ka$.

Making the usual approximation that $h_n^{(2)}(x)$ may

* The research reported in this paper has been sponsored largely by the AF Cambridge Research Center, Air Res. and Dev. Command, Bedford, Mass.

† National Bureau of Standards, Boulder, Colo.

be represented by Hankel functions of order $\frac{1}{3}$, the root determining equation may be written

$$\prod_{j=1,2} \left[\delta_j e^{i\pi/3} \frac{H_{2/3}^{(2)}[(1/3)(-2\tau_s)^{3/2}]}{H_{1/3}^{(2)}[(1/3)(-2\tau_s)^{3/2}]} + (-2\tau_s)^{-1/2} \right] = 0 \quad (97)$$

where

$$\delta_1 = \frac{-i}{(ka)^{1/3}\Delta_1}, \quad \delta_2 = \frac{-i}{(ka)^{1/3}\Delta_2} \text{ and } \tau_s = \frac{\nu - ka}{(ka)^{1/3}}.$$

The subscript s is affixed to τ in such a way that the mode of lowest attenuation is $s=0$ and $s=1, 2, 3, \dots$, corresponding to a progressive increase of attenuation. Since in most cases $|\Delta_1| \gg 1$ and $|\Delta_2| \ll 1$, only the roots corresponding to $j=2$ need be considered since those corresponding to $j=1$ are modes of relatively high attenuation. With this simplification, along with the replacement of the Legendre function by the leading term in its asymptotic expansion, one arrives at

$$U \cong 2U_0(-2\pi iX)^{1/2} \sum_{s=0}^{\infty} f_s(z) \frac{f_s(h)e^{-i\tau_s X}}{2\tau_s - 1/\delta_2^2} \quad (98)$$

with

$$f_s(z) = \left[\frac{X_s^2 - 2\tau_s}{-2\tau_s} \right]^{1/2} \frac{H_{1/3}^{(2)}[1/3(X_s^2 - 2\tau_s)^{3/2}]}{H_{1/3}^{(2)}[1/3(-2\tau_s)^{3/2}]} \quad (99)$$

where

$$U_0 = \frac{Ids}{4\pi i\omega\epsilon a\theta}, \quad h = b - a, \quad z = r - a$$

$$X = (ka)^{1/3}\theta \quad X_s = (ka)^{1/3}(2z/a)^{1/2} \quad (100)$$

and

$$X_h = (ka)^{1/3}(2h/a)^{1/2}.$$

The preceding equations are identical in form to those derived by Van der Pol and Bremmer [31] when modified to a spherical boundary of surface impedance equal to $120\pi\Delta_2$ [32]. The physical behavior of the problem is somewhat different though, because of the highly reactive nature of the surface. For example, the modal equation from which the modes of low attenuation are obtained is

$$\delta_2(3Z)^{1/3} = e^{i2\pi/3} H_{1/3}^{(2)}(Z)/H_{2/3}^{(2)}(Z) \quad (101)$$

where

$$Z = (1/3)(-2\tau)^{3/2}.$$

When $-2\pi < \arg Z < \pi$ and $|Z| \gg 1$, the right hand side of the preceding equation may be replaced by an asymptotic expansion; thus

$$\delta_2(3Z)^{1/3} \cong i - \frac{1}{6Z} - \frac{5}{144} \frac{i}{Z^2} + \frac{5}{648Z^3} \dots$$

which leads to the root

$$\tau' \cong \frac{1}{2\delta_2^2} \left(1 - \delta_2^3 - \frac{5}{8} \delta_2^6 \dots \right), \quad (102)$$

valid for $4\pi/3 > \arg \Delta_2 > \pi/3$. This yields a term in the residue series which behaves as

$$\frac{e^{-ika\theta} e^{-i\tau'X}}{(a\theta)^{1/2}} \exp \left[-ika\theta \left(1 - \frac{\Delta_2^2}{2} \right) \left(1 - \delta_2^3 - \frac{5}{8} \delta_2^6 \dots \right) \right] \cong \frac{e^{-ika\theta} e^{-i\tau'X}}{(a\theta)^{1/2}} \quad (103)$$

As δ_2 approaches zero or $ka \rightarrow \infty$ the field has the properties of a surface wave on a plane with a distribution of bosses.

This surface wave mode is the dominant one when the losses are small, since Δ_2 has a large imaginary part. The main influence of curvature in this case is to retard the phase, since δ_2^3 is approximately a negative real quantity.

In addition to the root τ' , there are other modes on the curved surface which can be obtained by perturbations of the solutions of

$$H_{2/3}^{(2)}(Z) = 0 \quad \text{as} \quad |\delta| \rightarrow \infty$$

and

$$H_{1/3}^{(1)}(Z) = 0 \quad \text{as} \quad |\delta| \rightarrow 0.$$

These are well known from the Van der Pol-Bremmer [31] theory, and for the two respective cases are:

$$\tau_s = \tau_{s,\infty} - \frac{1}{2\tau_{s,\infty}\delta_2} - \frac{1}{8\tau_{s,\infty}\delta_2^2} \dots \quad (s = 0, 1, 2, 3, \dots) \quad (104)$$

and

$$\tau_s = \tau_{s,0} - \delta - \frac{2}{3}\tau_{s,0}\delta^3 + \frac{1}{2}\delta^4 \dots \quad (s = 0, 1, 2, 3, \dots) \quad (105)$$

From the "tangent approximation" it follows that

$$\tau_{s,\infty} \cong \frac{1}{2} \{ 3\pi(s + \frac{1}{2}) \}^{2/3} e^{-i\pi/3}, \quad (s = 2, 3, 4, \dots) \quad (106)$$

and

$$\tau_{s,0} \cong \frac{1}{2} \{ 3\pi(s + 3/4) \}^{2/3} e^{-i\pi/3}, \quad (s = 2, 3, 4, \dots) \quad (107)$$

For $s=0$ and 1 the more accurate values given by

$$\tau_{0,\infty} = 0.808e^{-i\pi/3} \quad \tau_{0,0} = 1.856e^{-i\pi/3}$$

$$\tau_{1,\infty} = 2.577e^{-i\pi/3} \quad \tau_{1,0} = 3.245e^{-i\pi/3}$$

should be employed.

The influence of curvature can be displayed more clearly by using an alternative representation [32], [33] for the field which may be derived from the residue series quoted above. It reads

$$\frac{U}{2U_0} \cong G \left\{ F(p_0) - \frac{\delta_2^3}{2} [1 - i(\pi p_0)^{1/2} - (1 + 2p_0)F(p_0)] \right. \\ \left. + \delta_2^6 \left[1 - i(\pi p_0)^{1/2}(1 - p_0) - 2p_0 \right. \right. \\ \left. \left. + \frac{5}{6} p_0^2 + \left(\frac{p_0^2}{2} - 1 \right) F(p_0) \right] \right. \\ \left. + \text{terms in } \delta_2^9, \delta_2^{12}, \text{ etc.} \right\} \quad (108)$$

where

$$F(p_0) = 1 - i(\pi p_0)^{1/2} e^{-p_0} \operatorname{erfc}(i p_0^{1/2}), \quad (109)$$

$$p_0 = \frac{-ika\theta}{2} \Delta_2^2,$$

and

$$\delta_2^3 = \frac{1}{ka(i\Delta_2)^3}.$$

The "height-gain" factor G is given by

$$G \cong e^{+ik h \Delta_2} e^{+ik z \Delta_2} \quad (110)$$

in the present case.

It is thus evident that for small curvature (*i.e.*, $|\delta_2^3| \ll 1$), the field on the spherical surface is indeed very similar to that for a flat surface. The curvature correction factors are quite complicated but it may be noted that if $|p_0|$ is large, then

$$\frac{U}{2U_0} \cong GF(p_0) \left[1 + \delta_2^3 p_0 + \frac{\delta_2^6 p_0^2}{2} + \frac{\delta_2^9 p_0^3}{3!} + \dots \right] \\ \cong -Gi2(\pi p_0)^{1/2} e^{-p_0} e^{\delta_2^3 p_0}. \quad (111)$$

To the first order in δ_2^3 , this is the same as (103) above directly from the residue series.

PROPAGATION IN A ROUGH-WALLED WAVEGUIDE

Another application is the calculation of the fields in a parallel plate waveguide whose walls are rough in the sense described above. Choosing the usual cylindrical coordinate system (ρ, ϕ, z) , the waveguide region is bounded by perfectly conducting planes at $z=0$ and $z=H$, while each possesses a distribution of hemispherical bosses (see Fig. 7). The source is a vertical elec-

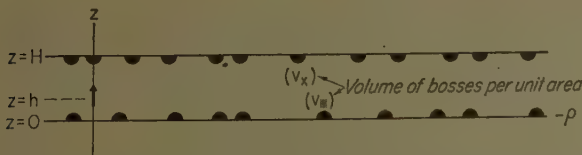


Fig. 7—A vertical dipole in a parallel-plate waveguide which has uniformly rough walls; in the examples considered in the text, $v_0=0$ and $v_x=v_y$.

tric dipole of strength Ids located at $\rho=0$ and $z=h$ where $0 < h < H$. The volume of the bosses per unit area is denoted v_0 for the boundary $z=0$, and v_x for the boundary $z=H$. The corresponding factors to account

for the electrical properties of the bosses are denoted Γ_0, Γ_0^* and Γ_x, Γ_x^* respectively.

As in the single-plane problem, the Hertz vector has only a z component which is proportional to a function ψ . It is defined such that the total magnetic field H_ϕ^t is given by

$$H_\phi^t = -b \frac{\partial}{\partial \rho} \psi^t \quad \text{where} \quad b = \frac{Ids}{4\pi}.$$

The boundary conditions written in terms of ψ are seen to be

$$\frac{\partial}{\partial z} \psi^t = \left(\alpha_0^* \frac{k^2}{2} - \alpha_0 k^2 - \alpha_0 \frac{\partial^2}{\partial z^2} \right) \psi^t \quad \text{at } z=0$$

and

$$\frac{\partial}{\partial z} \psi^t = - \left(\alpha_x^* \frac{k^2}{2} - \alpha_x k^2 - \alpha_x \frac{\partial^2}{\partial z^2} \right) \psi^t \quad \text{at } z=H$$

where $\alpha_j = 3v_j \Gamma_j$ and $\alpha_j^* = 3v_j^* \Gamma_j^*$ for $j=0$, or x . Using the same integral representation for the primary field as in Part I, the solution is readily obtained. It reads

$$\psi^t = \int_0^\infty F(\lambda) J_0(\lambda \rho) (\lambda/u) d\lambda \quad (112)$$

where

$$F(\lambda) = \frac{(e^{uz} + R_v^{(0)} e^{-uz})(e^{u(H-h)} + R_v^{(x)} e^{-u(H-h)})}{e^{uH}(1 - R_v^{(0)} R_v^{(x)} e^{-2uH})}, \quad (113)$$

and where

$$R_v^{(j)} = \frac{2u - (\alpha_j^* k^2 - 2\alpha_j \lambda^2)}{2u + (\alpha_j^* k^2 - 2\alpha_j \lambda^2)} \quad \text{for } j=0 \text{ or } x. \quad (114)$$

This may be rewritten in the form

$$R_v^{(j)} = -1 + \frac{2\Delta_{1j}}{K_j(C + \Delta_{1j})} - \frac{2\Delta_{2j}}{K_j(C + \Delta_{2j})} \quad (115)$$

where

$$\Delta_{1j} = \frac{i}{2k\alpha_j} (1 + K_j) \quad \text{and} \quad \Delta_{2j} = \frac{i}{2k\alpha_j} (1 - K_j),$$

with

$$K_j = \left[1 - 4 \left(1 - \frac{\alpha_j^*}{2\alpha_j} \right) (k\alpha_j)^2 \right]^{1/2}, \quad (116)$$

where C is now regarded as a new (dimensionless) variable related to u by $u = ikC$. Then, on changing the variable of integration to S where $\lambda = kS$ and using the identity $2J_0 = H_0^{(1)} + H_0^{(2)}$, it follows that

$$\psi^t = \frac{k}{2} \int_{-\infty}^{+\infty} F(C) H_0^{(2)}(kS\rho) \frac{SdS}{C} \quad (117)$$

where

$$F(C) = \frac{(e^{ikCz} + R_v^{(0)} e^{-ikCz})(e^{ikC(H-h)} + R_v^{(x)} e^{-ikC(H-h)})}{e^{ikCH}(1 - R_v^{(0)} R_v^{(x)} e^{-2ikCH})}. \quad (118)$$

The integration contour is along the real axis in the S plane. If there are no losses, the contour should be indented upward at the branch point $S = +1$ and downward at the branch point $S = -1$. There are apparently no further branch points so the integral may be evaluated by summing the residues of the poles in the lower half-plane of the two-sheeted Riemann surface. This leads to

$$\psi^i = -i\pi k \sum_n \left[\frac{1}{\frac{\partial}{\partial C} \frac{1}{F(C)}} \right]_{C=C_n} H_0^{(2)}(kS_n\rho) \quad (119)$$

where C_n is the solution of the modal equation

$$R_v^{(0)} R_v^{(x)} e^{-2ikC_n H} = e^{-2\pi i n} \quad (120)$$

for $n = 0, 1, 2, 3, \dots$ and $S_n^2 = (1 - C_n^2)^{1/2}$. The solution may be written more explicitly in the form

$$\psi^i = \frac{\pi}{h} \sum_n f_n(h) f_n(z) \delta_n H_0^{(2)}(kS_n\rho) \quad (121)$$

where

$$\delta_n = \frac{1}{\left[1 + i \frac{\partial [R_v^{(0)} R_v^{(x)}] / \partial C}{2kh R_v^{(0)} R_v^{(x)}} \right]_{C=C_n}}, \quad (122)$$

$$f_n(z) = \left[\frac{e^{ikC_n z} + R_v^{(0)} e^{-ikC_n z}}{2(R_v^{(0)})^{1/2}} \right]_{C=C_n}, \quad (123)$$

and $f_n(h)$ has exactly the same form as $f_n(z)$. In the limiting case of perfectly conducting walls (*i.e.*, absence of bosses), the above reduces to the well known form

$$\psi^i = \frac{\pi}{h} \sum_n \epsilon_n \cos(kC_n h) \cos(kC_n z) H_0^{(2)}(kS_n\rho) \quad (124)$$

where $\epsilon_0 = 1$, $\epsilon_n = 2$, ($n = 1, 2, 3, \dots$) and where $C_n = (\pi n / kH)$.

The properties of the modes are now investigated. To simplify the algebra, one wall of the guide is taken to be perfectly conducting and smooth. Thus $\Delta_{10} = \infty$ and $\Delta_{20} = 0$ or $R_v^{(1)} = 1$ and $\Delta_{1x} = \Delta_1$ and $\Delta_{2x} = \Delta_2$ where

$$\Delta_{1,2} = \frac{i}{6kv} \left[1 \pm \sqrt{1 - 4 \left(1 - \frac{\Gamma^*}{2} \right) (3kv)^2} \right]$$

for non-magnetic bosses. The modal equation then becomes

$$-\left[\frac{C_n - \Delta_1}{C_n + \Delta_1} \right] \left[\frac{C_n - \Delta_2}{C_n + \Delta_2} \right] e^{-i2kC_n H} = e^{-i2\pi n}. \quad (125)$$

In the general case, it is apparently necessary to solve this equation by numerical means. Two limiting cases, however, can be readily treated. Remembering that for small roughness $|\Delta_1| \gg 1$ and $|\Delta_2| \ll 1$, it is assumed that for the first limiting case $|C_n| \ll |\Delta_1|$ and $|C_n| \gg |\Delta_2|$.

The approximate form of the mode equation then becomes

$$(C_n / \Delta_1) + (\Delta_2 / C_n) + ikC_n H = i\pi n \quad (126)$$

which may be solved for C_n to yield

$$S_n = \left\{ 1 - \left(\frac{\pi n}{kH} \right)^2 \cdot \frac{1}{4} \left[1 \pm \left(1 + \frac{4i\Delta_2 kH}{(\pi n)^2} \right)^{1/2} \right]^2 \right\}^{1/2} \quad (127)$$

where

$$\bar{H} = H \left(1 + \frac{1}{ikH\Delta_1} \right).$$

The positive sign before the radical is chosen since C_n is to reduce to $(\pi n / kH)$ as $\Delta_2 \rightarrow 0$ and $\Delta_1 \rightarrow \infty$. For $n = 0$, this simplifies to

$$S_0 = \left[1 - i \frac{\Delta_2}{kH} \right]^{1/2}.$$

For $n > 0$, (127) may be simplified (if $|\Delta_2 kH| \ll 1$ and $|\Delta_1 kH| \gg 1$) to

$$S_n \cong \left[1 - \left(\frac{\pi n}{kH} \right)^2 - i \frac{2\Delta_2}{kH} \right]^{1/2} \quad \text{for } n = 1, 2, 3, \dots \quad (128)$$

which to a first order does not depend on Δ_1 . If in addition to $kH|\Delta_2| \ll 1$ also

$$\frac{|\Delta_2|}{kH} \ll 1 - \left(\frac{\pi n}{kH} \right)^2,$$

it is possible to write

$$\text{Re } S_n \cong \left[1 - \left(\frac{\pi n}{kH} \right)^2 \right]^{1/2} + \frac{\epsilon_n \text{Im. } \Delta_2}{2kH} \left[1 - \left(\frac{\pi n}{kH} \right)^2 \right]^{-1/2} \quad (129)$$

$$\text{Im } S_n \cong - \frac{\epsilon_n \text{Re. } \Delta_2}{2kH} \left[1 - \left(\frac{\pi n}{kH} \right)^2 \right]^{-1/2} \quad (130)$$

where $\epsilon_0 = 1$ and $\epsilon_n = 2$ for $n = 1, 2, 3, \dots$. The latter two expressions are appropriate for small roughness and for modes which are not near cutoff. It is interesting to note that the phase velocity (proportional to $1/\text{Re. } S_n$) is decreased by the presence of roughness since $\text{Im. } \Delta_2 > 0$. As expected, the finite conductivity of the bosses produces attenuation.

Another approximate solution can be obtained under the conditions

$$|C_n| \ll |\Delta_1| \quad \text{and} \quad |C_n| \ll |\Delta_2|$$

which leads to

$$(C_n / \Delta_1) + (C_n / \Delta_2) + ikC_n H \cong i\pi(n - 1/2). \quad (131)$$

Thus

$$C_n \cong \frac{\pi(n - \frac{1}{2})}{kH - i\left(\frac{1}{\Delta_1} + \frac{1}{\Delta_2}\right)} \quad (132)$$

and

$$S_n = (1 - C_n^2)^{1/2}.$$

If $(1/\Delta_1 + 1/\Delta_2) \ll kH$, one may write

$$S_n = \left[1 - C_{0,n}^2 - \frac{i2}{kH} \left(\frac{1}{\Delta_1} + \frac{1}{\Delta_2}\right) C_{0,n}^2\right]^{1/2}; \quad (133)$$

and, if in addition

$$\left(\frac{1}{\Delta_2} + \frac{1}{\Delta_1}\right) \frac{1}{kH} C_{0,n}^2 \ll 1 - C_{0,n}^2,$$

it follows that

$$\begin{aligned} \text{Re. } S_n &\cong [1 - C_{0,n}^2]^{1/2} \\ &+ \text{Im.} \left(\frac{1}{\Delta_1} + \frac{1}{\Delta_2}\right) C_{0,n}^2 [1 - C_{0,n}^2]^{-1/2} \frac{1}{kH} \end{aligned} \quad (134)$$

and

$$\text{Im. } S_n = -\text{Re.} \left(\frac{1}{\Delta_1} + \frac{1}{\Delta_2}\right) C_{0,n}^2 [1 - C_{0,n}^2]^{-1/2} \frac{1}{kH} \quad (135)$$

where

$$C_{0,n} = \frac{\pi(n - \frac{1}{2})}{kH}.$$

These latter expressions are appropriate for low order modes when the roughness is not too small. It is seen that the phase velocity to a first order corresponds to a waveguide whose wall is acting as a perfect magnetic conductor. The second term on the right hand side of (134) is a negative correction to $\text{Re. } S_n$ and thus tends to increase the phase velocity. For finitely conducting bosses the quantity $\text{Im. } S_n$ is negative and leads to attenuation. It is interesting to note that the low order modes and large waveguides are associated with small attenuation.

The excitation of the waveguide by a vertical magnetic dipole can be treated in the same fashion as for the vertical electrical dipole. The field has the same form as in the electric dipole case and is obtained by simply replacing $R_v^{(0)}$ and $R_v^{(x)}$ by $R_h^{(0)}$ and $R_h^{(x)}$, respectively, where the latter reflection coefficients are of a type given by (71). Again restricting discussion to a waveguide with one smooth and one rough wall, the relevant modal equation becomes

$$\frac{2 + ik\alpha^* C_m}{2 - ik\alpha^* C_m} e^{-i2kHC_m} = e^{-i2\pi m} \quad (136)$$

for $m = 1, 2, 3, \dots$

For $|k\alpha^* C_m| \ll 1$, it easily follows that

$$C_m \cong \frac{\pi m}{(kH - k\alpha^*/2)} \quad (137)$$

and

$$S_m = (1 - C_m^2)^{1/2}.$$

If $|\alpha^*| \ll H$

$$S_m \cong \left[1 - \left(\frac{\pi m}{kH}\right)^2 - \frac{\alpha^*}{H} \left(\frac{\pi m}{kH}\right)^2\right]^{1/2}, \quad (138)$$

and if in addition

$$\frac{\alpha^*}{H} \left(\frac{\pi m}{kH}\right)^2 \ll 1 - \left(\frac{\pi m}{kH}\right)^2$$

it follows that

$$\begin{aligned} \text{Re. } S_m &= \left[1 - \left(\frac{\pi m}{kH}\right)^2\right]^{1/2} \\ &- \frac{\text{Re. } \alpha^*}{2H} \left(\frac{\pi m}{kH}\right)^2 \left[1 - \left(\frac{\pi m}{kH}\right)^2\right]^{-1/2} \end{aligned} \quad (139)$$

and

$$\text{Im. } S_m = -\frac{\text{Im. } \alpha^*}{2H} \left(\frac{\pi m}{kH}\right)^2 \left[1 - \left(\frac{\pi m}{kH}\right)^2\right]^{-1/2}. \quad (140)$$

The behavior of the phase velocity and the attenuation factor for horizontal polarization has a very similar behavior to that of a moderately rough walled guide at vertical polarization for low order modes. In general, however, if everything is constant the influence of wall roughness for horizontal polarization (*i.e.* T.E. waves) is much smaller than for vertical polarization (*i.e.* T.M. waves).

INFLUENCE OF CURVATURE IN A ROUGH-WALLED WAVEGUIDE

The propagation between two concentric curved surfaces which may both have uniform roughness can be treated in a further extension (see Fig. 8). In the case of spherical surfaces it is appropriate to employ a mode expansion in terms of spherical wave functions. The boundary conditions are applied at both surfaces ($r = a$ and $r = c$, respectively). In the case when the lower

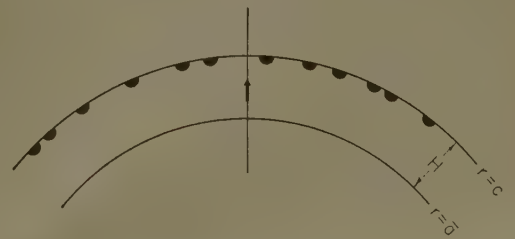


Fig. 8—A vertical dipole between spherical concentric surfaces; the upper is uniformly rough and the lower is smooth.

surface is smooth and the upper surface only is rough, the modal equation for T.M. modes is

$$\prod_{j=1,2} \left[\frac{\ln' [kch_v^{(2)}(kc)] + i\Delta_j}{\ln' [kch_v^{(1)}(kc)] + i\Delta_j} \right] \frac{h_v^{(1)}(ka)h_v^{(2)}(kc)}{h_v^{(2)}(ka)h_v^{(1)}(kc)} + 1 = 0. \quad (141)$$

This is the spherical analog to (125) for the parallel-plate guide. On writing $v + \frac{1}{2} = kaS_n$, using the second-order approximations [34] to the spherical Hankel functions, the preceding equation may be written

$$-\left(\frac{C_n' - \Delta_1}{C_n' + \Delta_1}\right)\left(\frac{C_n' - \Delta_2}{C_n' + \Delta_2}\right) \cdot \exp \left[-i2k \int_0^H \left(C_n^2 + \frac{2z}{a} S_n^2 \right)^{1/2} dz \right] = e^{-i2\pi n} \quad (n = 0, 1, 2, \dots) \quad (142)$$

where $H = c - a$ and

$$C_n' \cong \left[C_n^2 + \frac{2H}{a} S_n^2 \right]^{1/2}.$$

The preceding modal equation is valid if $\text{Re. } S_n < 1$ and $\text{Re. } C_n \gg (2/ka)^{1/3}$. When $(2H/a) \ll |C_n^2|$ the preceding form of the modal equation reduces to (125) as it should.

The numerical treatment of the modal equation for concentric spherical boundaries is more involved, although if the curvature is small a perturbation method may be employed. This is illustrated for the case of moderate roughness such that $C_n' \ll \Delta_2$. Then to a first order in (H/a) the modal equation may be written

$$C_n \left(\frac{1}{\Delta_1} + \frac{1}{\Delta_2} \right) + ikHC_n + iG(C_n) = i\pi(n - \frac{1}{2}) \quad (143)$$

where

$$G(C_n) \cong \frac{H}{a} \frac{S_n^2}{C_n^2} \left[\frac{kHC_n}{2} - \frac{C_n}{\Delta} \right]. \quad (144)$$

When G is replaced by zero the zero-order solution corresponding to the parallel-plate case is obtained. This is given by (132). This value for C_n is now inserted into $G(C_n)$ which is denoted G_0 . Eq. (143) is now resolved to obtain the following first-order solution:

$$S_n \cong \left[1 - C_{0,n}^2 \left(1 - \frac{2G_0}{\pi(n - \frac{1}{2})} + \frac{2i}{\Delta kH} \right) \right]^{1/2} \quad (145)$$

where

$$C_{0,n} \cong \frac{\pi(n - \frac{1}{2})}{kH}$$

$$G_0 \cong \frac{H}{a} \frac{S_{0,n}^2 kH}{2C_{0,n}} - \frac{3i}{2\Delta} \frac{H}{a} \frac{S_{0,n}^2}{C_{0,n}}, \text{ and}$$

$$S_{0,n} \cong (1 - C_{0,n}^2)^{1/2}.$$

When

$$\left(\frac{1}{\Delta_1 + \Delta_2} \right) \frac{C_{0,n}^2}{kH} \ll S_{0,n},$$

it is possible to write

$$\text{Re. } S_n \cong S_{0,n} \left(1 + \frac{H}{2a} \right) + \text{Im.} \left(\frac{1}{\Delta_1} + \frac{1}{\Delta_2} \right) \frac{C_{0,n}^2}{S_{0,n} kH} \quad (146)$$

and

$$\text{Im. } S_n \cong -\text{Re.} \left(\frac{1}{\Delta_1} + \frac{1}{\Delta_2} \right) \frac{C_{0,n}^2}{S_{0,n} kH} \cdot \left(1 + \frac{3H}{2a} \frac{S_{0,n}^2}{C_{0,n}^2} \right). \quad (147)$$

These equations show that the influence of curvature, at least for the present example, is to retard the phase and to increase the attenuation relative to a parallel plate guide. Models, such as the one described above, explain some of the observed characteristics of VLF terrestrial radio waves which have propagated to great distances [35].

ACKNOWLEDGMENT

The author would like to thank Mrs. A. Fails for her assistance in preparing the manuscript.

BIBLIOGRAPHY

- [31] H. Bremmer, "Terrestrial Radio Waves," Elsevier Publishing Co., New York, N. Y., and Amsterdam, Holland, pp. 42-45; 1949.
- [32] J. R. Wait, "Radiation from a vertical antenna on a curved stratified ground," *J. Res. Natl. Bur. Standards*, vol. 56, pp. 237-244; April, 1956.
- [33] H. Bremmer, "Applications of operational calculus to ground wave propagation," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-6, pp. 267-273; July, 1958.
- [34] A. N. Sommerfeld, "Partial Differential Equations," Academic Press, New York, N. Y., pp. 118-122; 1949.
- [35] J. R. Wait, "Theory of the terrestrial propagation of V.L.F. radio waves," to be published as a Natl. Bur. of Standards Monograph.

Some Theoretical Results for Surface Wave Launchers

J. BROWN†

Summary—The design of surface wave launchers is examined and it is shown that the efficiency of a launcher of finite size can approach arbitrarily near to 100 per cent. For a given size of launcher, there is however a limit to the efficiency above which an increase is obtained only at the expense of frequency bandwidth. Launchers with efficiencies above this limit are classed as "super-efficient" by analogy with "super-gain" aeriels. A general solution for an end-fire launcher is also given.

I. INTRODUCTION

ALTHOUGH a surface wave corresponds to a solution of Maxwell's equations which is capable of existing independently of any other fields, it is now well known that in situations of practical interest the surface wave is accompanied by a radiation field. The principal reason for this is that it is impossible to launch a pure surface wave, uncontaminated by radiation, unless the field appropriate to the launcher is established over a plane of infinite extent in at least one dimension. Practical launchers are necessarily restricted to a finite area and the surface wave is then accompanied by a radiation field. The concept of launching efficiency¹ has been introduced to give a measure of the effectiveness of a launcher and is defined as the percentage of the total power, leaving the launcher, which is associated with the surface wave. Launching efficiency is thus a quantity analogous to the power gain of an aerial, as it indicates the effectiveness of the system in concentrating the power in some desired manner. It has been shown that there is no limit to the power gain which can be achieved by an aerial of a given aperture size. This suggests that, in principle, a launcher of prescribed size can be designed to have a launching efficiency arbitrarily near to the maximum possible value of 100 per cent. This problem is discussed in Section II. In Section III a general solution is given for a launcher similar in arrangement to an end-fire aerial.

II. RADIATION FIELD EXCITED BY LAUNCHER

Summary of General Results

The arrangement considered is the general two-dimensional problem first investigated by Cullen.¹ A surface of reactance X is positioned in the half plane $y=0$, $x \geq 0$; and is excited by a launcher which has an aperture lying in the half plane $x=0$, $y \geq 0$. The only non-zero field components are E_x , E_y , and H_z . E_y is prescribed over

the launcher aperture as

$$E_y(0, y) = f(y); y \geq 0.$$

The general solution obtained by Cullen can be written as

$$H_z(x, y) = \frac{\omega\epsilon}{2\pi} \int_{-\infty}^{\infty} \left[P(\xi) + \frac{\xi - ju}{\xi + ju} P(-\xi) \right] \exp[-j\xi y - j\sqrt{k^2 - \xi^2} x] \frac{d\xi}{\sqrt{k^2 - \xi^2}} + \frac{2\omega\epsilon u}{\beta} \exp[-uy - j\beta x] P(ju) \quad (1)$$

where

$$u = \omega\epsilon X \quad (2)$$

$$\beta = \sqrt{k^2 + u^2} \quad (3)$$

$$P(\xi) = \int_0^{\infty} f(y) e^{i\xi y} dy. \quad (4)$$

The last term in (1) is the surface wave, and the integral, which is identical in form to the expression obtained by Booker and Clemmow² for the radiation field over an imperfect conductor, also gives the radiation field in this case.

The radiation pattern is usually investigated at large distances from the source, it then being possible to evaluate the integral to a sufficient accuracy by the first term of the stationary phase method. This gives, for the radiation field, $H_z^R(x, y)$:

$$H_z^R(x, y) = \frac{\omega\epsilon \exp[-j(kr + \pi/4)]}{\sqrt{2\pi kr}} \cdot \left[P(k \sin \theta) + \left(\frac{jk \sin \theta + u}{jk \sin \theta - u} \right) P(-k \sin \theta) \right] \quad (5)$$

where

$$x = r \cos \theta, \quad y = r \sin \theta.$$

The total radiated power can be obtained by evaluating the integral of $|H_z^R(x, y)|^2$ over the range of angles 0 to $\pi/2$, while the power transmitted by the surface wave is calculated from the amplitude as given by (1). The launching efficiency follows.

† University College, Gower St., London, England.

¹ A. L. Cullen, "The Excitation of Plane Surface Waves," Monograph No. 93, *Proc. IEE*, vol. 101, pt. IV, pp. 225-234; October, 1954.

² H. G. Booker and P. C. Clemmow, "Sommerfeld theory and straight edge diffraction theory," *Proc. IEE*, vol. 97, pt. III, pp. 18-27; January, 1950.

Design of Launchers with Maximum Efficiency

The problem investigated here is the determination of the maximum launching efficiency if the launcher aperture is restricted to the strip $0 \leq y \leq a$ in the plane $x=0$. In other words, we restrict the function $f(y)$ to be zero for $y > a$, and we wish to calculate the shape of the aperture distribution which will give the highest launching efficiency. The line of attack is to consider a set of distributions $f(y)$, all of which give the same surface wave amplitude, and then to determine which member of the set gives the least power in the radiation field. The set of distributions considered will be taken of the form

$$f(y) = f_s(y) + \sum_n a_n f_n(y), \quad (6)$$

where $f_s(y)$ is a function which gives the required surface wave amplitude, and $f_1(y), f_2(y), \dots$ are functions, all of which give zero surface wave amplitude; *i.e.*, they represent distributions which give rise to a pure radiation field only. It will be shown below that a set of such functions do exist, and a convenient choice of the form of these functions will be made. Our problem is now one stage more definite: to choose the amplitudes a_n , in (6) in such a way that the radiation field caused by the summation cancels, as nearly as possible, the radiation field caused by $f_s(y)$. It is therefore obvious that the choice of $f_s(y)$ is unimportant, since any change in the radiation pattern caused by $f_s(y)$ can be compensated by appropriate changes in the amplitudes a_n . A convenient choice for $f_s(y)$ is the "chopped surface wave distribution,"

$$f_s(y) = \exp(-uy), \quad 0 \leq y \leq a. \quad (7)$$

Detailed expressions for the surface wave amplitude and the radiation pattern appropriate to this distribution follow readily from the general results already outlined and have been given by Cullen.

Choice of Aperture Functions

The problem of finding the aperture distribution in an aerial which will produce a prescribed radiation pattern has been discussed by Woodward and Lawson.³ We wish to extend their results to the present case in which the radiation from the aperture is modified by the presence of the reactive surface. We therefore need to find a set of functions $f_n(y)$ which will play the same part as do the functions $\exp(2n\pi jy/a)$ in the Woodward and Lawson analysis. The functions $f_n(y)$ have already been stated to be such that they do not give rise to any surface wave field. The condition for this follows directly

from (1) and (2) as

$$\int_0^a f_n(y) \exp(uy) dy = 0, \quad (8)$$

and it is obvious that a wide range of possibilities exist. A further necessary condition is that the functions should be orthogonal over the aperture; *i.e.*, that

$$\int_0^a f_n(y) f_m(y) dy = 0, \quad (9)$$

unless $m=n$. Unless this condition is satisfied, the expression of an aperture distribution in the form of (6) will not be unique.

A suitable choice of functions to satisfy (9) is

$$\sin(n\pi y/a): n=1, 2, \dots$$

but these do not satisfy (8). We therefore consider the set $\sin(n\pi y/a + \phi_n)$, and select ϕ_n so that (8) is satisfied; *i.e.*,

$$\int_0^a \sin(n\pi y/a + \phi_n) e^{-uy} dy = 0$$

which leads to

$$\tan \phi_n = -n\pi/au. \quad (10)$$

With this choice of ϕ_n , (9) remains valid.

It may be considered curious that the same choice of values for ϕ_n should satisfy both (8) and (9) for there is no guarantee that this should be so. An interesting sidelight on this point, which in fact led to the above choice of functions, is given by asking what solutions satisfy Maxwell's equations and the reactance boundary condition, and also have the property that the ratio of the transverse electric and magnetic fields is constant over the aperture. The functions

$$\exp -jk(x \cos \theta + y \sin \theta)$$

have this property in free space, and the Woodward and Lawson aperture functions are obtained by selecting a set of values of $\sin \theta$ to ensure that the corresponding functions are orthogonal. For the surface wave problem considered here, the functions $\sin[p y + \phi(p)] \exp -jx\sqrt{k^2 - p^2}$ have the required property if $\tan \phi(p) = -p/u$. The $f_n(y)$ above are of the form $\sin[p y + \phi(p)]$, the possible values of p being selected to give the orthogonality condition over the aperture.

An alternative derivation of the functions $f_n(y)$ could be based on finding the orthogonal set which satisfies both the wave equation and the boundary conditions.⁴

The next step is to determine the radiation fields corresponding to each function $f_n(y) = \sin(n\pi y/a + \theta_n)$.

³ P. M. Woodward and J. D. Lawson, "The theoretical precision with which an arbitrary radiation pattern may be obtained from a source of finite size," *Proc. IEE*, vol. 95, pt. III, pp. 363-370; September, 1948.

⁴ A. Sommerfeld, "Partial Differential Equations," Academic Press, New York, N. Y.; 1949.

From (4),

$$P_n(\zeta) = \int_0^a f_n(y) \exp(j\zeta y) dy$$

$$= \{(-)^n \exp(j\zeta a) - 1\} \left[\frac{\exp(j\phi_n)}{\zeta + n\pi/a} - \frac{\exp(-j\phi_n)}{\zeta - n\pi/a} \right]. \quad (11)$$

From (10),

$$\exp(2j\phi_n) = \frac{u - jn\pi/a}{u + jn\pi/a} \quad (12)$$

so that (11) can be rewritten as

$$P_n(\zeta) = -\frac{2n\pi \exp(-j\phi_n)(u + j\zeta) \{(-)^n \exp(j\zeta a) - 1\}}{(au + jn\pi)(\zeta^2 - n^2\pi^2/a^2)}. \quad (13)$$

The function of interest for the radiation field is

$$P_n(\zeta) + \frac{\zeta - ju}{\zeta + ju} P_n(-\zeta),$$

and this is found to be

$$\frac{4n\pi j(u + j\zeta)(-)^{n-1} \exp(-j\phi_n) \sin(\zeta a)}{(au + jn\pi)(\zeta^2 - n^2\pi^2/a^2)}.$$

Comparison with (5) shows that we can write the radiation field appropriate to $f_n(y)$ as

$$H_{z,n}^R = K_n \frac{\exp(-jkr)(au + jka \sin \theta) \sin(ka \sin \theta)}{\sqrt{kr}(k^2 a^2 \sin^2 \theta - n^2 \pi^2)}. \quad (14)$$

This expression has a marked similarity to the corresponding one in the Woodward and Lawson analysis in that $H_{z,n}^R$ vanishes when $ka \sin \theta = m\pi$, m being any integer except n . K_n is a constant, independent of r or θ , but depending on n , and is given by

$$K_n = \frac{\sqrt{8\pi j \omega \epsilon a} (-)^{n-1} \exp j\left(\frac{\pi}{4} - \phi_n\right)}{(au + jn\pi)}.$$

The radiation field created by the aperture distribution of (6) is

$$H_z^R = \frac{\exp(-jkr)}{\sqrt{Kr}} \left[C \frac{\sin(ka \sin \theta)}{ua - jka \sin \theta} + \sum_n a_n K_n \frac{(ua + jka \sin \theta) \sin(ka \sin \theta)}{(k^2 a^2 \sin^2 \theta - n^2 \pi^2)} \right] \quad (15)$$

where

$$C = -\sqrt{\frac{2}{\pi}} \omega \epsilon a \exp(-ua + j\pi/4). \quad (16)$$

Maximum launching efficiency corresponds to minimum total power in the radiation field. The aperture distribution for this is found by selecting the constants a_n to make

$$\int_0^{\pi/2} \left| \left[C + \sum_n a_n K_n \frac{(u^2 a^2 + k^2 a^2 \sin^2 \theta)}{k^2 a^2 \sin^2 \theta - n^2 \pi^2} \right] \frac{\sin(ka \sin \theta)}{ua - jka \sin \theta} \right|^2 d\theta \quad (17)$$

a minimum.

In this expression, C , a_n , and K_n are complex constants. Consideration of the real and imaginary parts of the integrand shows that for a minimum value the phases of C and each product $a_n K_n$ must be equal; i.e., if we write

$$C = |C| e^{j\alpha}, \quad (18)$$

then, for minimum, $a_n K_n = c_n e^{j\alpha}$ where the c_n are all real.

It may be noted that since the K_n are complex, the amplitudes a_n will have different phases, and the optimum aperture distribution will not, therefore, have a constant phase. The expression to be minimized can be rewritten as

$$\int_0^{\pi/2} \left[|C| + \sum_n \frac{c_n (k^2 a^2 \sin^2 \theta + u^2 a^2)}{k^2 a^2 \sin^2 \theta - n^2 \pi^2} \right] \frac{\sin^2(ka \sin \theta)}{k^2 a^2 \sin^2 \theta + u^2 a^2} d\theta. \quad (19)$$

The normal procedure of differentiating with respect to the quantities c_n , which have to be determined, leads to a set of equations which are extremely badly conditioned. It would thus require enormous computational labor to extract any numerical results. It is, however, scarcely necessary to perform calculations to determine the effect of modifying the distribution from the chopped surface wave value. We note, firstly, that the pattern corresponding to $f_n(y)$ has its largest maximum at $ka \sin \theta = n\pi$. If $ka < \pi$, this maximum will lie in the range of values of $\sin \theta$ corresponding to imaginary values. This means that in addition to the radiation field with which we are concerned, there will be a strong reactive field in the vicinity of the launcher aperture. Any attempt to increase the launching efficiency above the value appropriate to the chopped surface wave will, therefore, inevitably increase the reactive field near the aperture. Even if ka exceeds π , it is still true that the launching efficiency can be increased only at the expense of an increase in the storage field. The reason for this is that the terms in (19) whose largest maxima do lie in

the region of real angles have these maxima at angles where the radiation from the chopped surface distribution is already zero. Therefore, there is no point in using such terms as they only serve to increase the radiation field. Cancellation must be effected by the terms whose largest maxima lie in the $\sin \theta$ range corresponding to reactive fields.

Since there is an infinite number of constants available in (19) to be adjusted, it is virtually certain that the integral can be reduced to an arbitrarily small value. Therefore, it follows that in principle, the radiation power can be reduced to an arbitrarily small value; *i.e.*, the launching efficiency can be made as close to 100 per cent as we please. The reactive fields must, however, increase very rapidly when large values of n are used, since the effectiveness of a unit amplitude term in the radiation field region can be seen to be proportional to $1/n^2$. The detailed study of the effects of the reactive fields can be carried through in the way described by Harrington⁵ for the corresponding aerial problem.

This situation is closely parallel to that for aerials for which the power gain of an aperture of any arbitrary size can be made indefinitely large. There is, however, for an aperture of a given size, a limit to the gain, and any increase above this limit is obtained only at the expense of increasing the reactive field. The limit for a plane aperture corresponds to a field distribution which has constant phase and amplitude over the aperture. This aperture distribution is identical to the distribution of a plane wave which is the limiting form of the radiation pattern when the gain goes to infinity. Similarly, we can see from (19) that the chopped surface wave distribution gives a similar limit on the launching efficiency of an aperture of given size. Any change from this distribution will either increase the radiated power, thus decreasing the aperture efficiency, or it will lead to a marked increase in the reactive field. The presence of a reactive field near the aperture is objectionable in that it leads to a reactive component in the impedance presented to the feeder and this restricts the frequency range over which satisfactory power transfer from the feeder to the launcher can be achieved.

The term "super-gain" is applied to aerials in which the gain exceeds that for the aperture distribution equivalent to the plane wave. A similar term, "super-efficient," can be used to describe launchers whose efficiency exceeds that of the chopped surface wave distribution. Super-efficient launchers are inevitably narrow band devices. The slot launcher described by Cullen is an example of a super-efficient launcher, and the difficulties encountered in satisfactorily matching such a system have been described elsewhere.⁶

⁵ R. F. Harrington, "On the gain and beamwidth of directional antennas," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-6, pp. 219-225; July, 1958.

⁶ J. Brown and H. S. Stachera, "Annular slot launchers for single-conductor transmission lines," to be published in *Proc. IEE*, vol. 106, suppl. to pt. B.

III. END-FIRE LAUNCHERS

An attractive form of launcher is the analog of an end-fire aerial. Such a launcher could, for example, consist of a coaxial line contained within a cylinder over which a surface reactance was provided. The corresponding two-dimensional arrangement has a reactive surface over the half-plane $y=0, x \geq 0$ as before, while a prescribed field, for example E_x , is defined over the half-plane $y=0, x \leq 0$. The problem involved in calculating the launching efficiency is of the type amenable to solution by Wiener-Hopf methods.

As in the other problem considered, the only non-zero field components are E_x, E_y , and H_z , and the conditions of the problem require that

$$j\omega\epsilon E_x = -uH_z \text{ for } y=0, x \geq 0 \quad (20)$$

where u is defined in terms of the surface reactance by (2).

$$E_x = f(x) \text{ for } y=0, x \leq 0, \quad (21)$$

$f(x)$ being the prescribed aperture distribution created by the launcher. The general solution for H_z in the upper half plane $y \geq 0$ can be written as

$$H_z = \frac{\omega\epsilon}{2\pi} \int_{-\infty}^{\infty} P(\zeta) \exp[-j\zeta x - j\sqrt{k^2 - \zeta^2}y] d\zeta. \quad (22)$$

$\sqrt{k^2 - \zeta^2}$ is defined as the branch which has the value k when $\zeta=0$. It will be assumed that k has a small negative imaginary part corresponding to a slight attenuation in the region $y \geq 0$.

The conditions (20) and (21) give

$$\int_{-\infty}^{\infty} [u - j\sqrt{k^2 - \zeta^2}] P(\zeta) \exp(-j\zeta x) d\zeta = 0, \quad x \geq 0 \quad (23)$$

$$\int_{-\infty}^{\infty} \sqrt{k^2 - \zeta^2} P(\zeta) \exp(-j\zeta x) d\zeta = -2\pi f(x), \quad x \leq 0. \quad (24)$$

From (23), we have

$$[u - j\sqrt{k^2 - \zeta^2}] P(\zeta) = L(\zeta) \quad (25)$$

where $L(\zeta)$ is a function of ζ which is regular in the lower half plane, $\text{Im } \zeta < -\text{Im } k$.

From (24)

$$\sqrt{k^2 - \zeta^2} P(\zeta) = F(\zeta) + U(\zeta) \quad (26)$$

where $U(\zeta)$ is regular in the upper half plane, $\text{Im } \zeta > \text{Im } k$ and

$$F(\zeta) = \int_{-\infty}^0 f(x) \exp(j\zeta x) dx. \quad (27)$$

If

$$f(x) \rightarrow C \exp(\alpha x) \text{ as } x \rightarrow -\infty, \quad (28)$$

$F(\xi)$ will be regular for $\text{Im } \xi < \alpha$. From (27) and (28)

$$\frac{\sqrt{k^2 - \xi^2}}{u - j\sqrt{k^2 - \xi^2}} L(\xi) = F(\xi) + U(\xi). \quad (29)$$

Further, the function

$$\frac{\sqrt{k^2 - \xi^2}}{u - j\sqrt{k^2 - \xi^2}}$$

can be written

$$\frac{\sigma_-(\xi)}{\sigma_+(\xi)}$$

where $\sigma_-(\xi)$ is regular and free from zeros for $\text{Im } \xi < -\text{Im } k$ and $\sigma_+(\xi)$ is regular and free from zeros for $\text{Im } \xi > \text{Im } k$. All the functions in (29) are now regular for $\text{Im } k < \text{Im } \xi < -\text{Im } k$ so that the usual Wiener-Hopf technique should be applicable. A difficulty arises, however, in that $F(\xi)$ is not expressed in a form where its singularities are explicit. This complicates the procedure for separating the equation into two parts which are, respectively, regular in the upper and lower half planes.

Suppose then that

$$f(x) = \exp[-j\beta x + \alpha x] \quad (30)$$

where β is a real constant.

In this case,

$$F(\xi) = \frac{1}{(j\xi - j\beta + \alpha)} \quad (31)$$

and (29) can now be split without any difficulty giving

$$L(\xi) = -\frac{\sigma_+(\beta + j\alpha)}{(j\xi - j\beta + \alpha)} \quad (32)$$

when the solution for the field is

$$H_z = \frac{\omega\epsilon}{2\pi} \int_{-\infty}^{\infty} \frac{\sigma_+(\beta + j\alpha) \exp(-j\xi x - j\sqrt{k^2 - \xi^2} y)}{(j\xi - j\beta + \alpha)(u - j\sqrt{k^2 - \xi^2})} d\xi. \quad (33)$$

This can be expressed as a surface wave field, arising from the pole at $j\sqrt{k^2 - \xi^2} = u$, and a radiation field, as in the problem of the excitation of radiation at a discontinuity in surface reactance discussed by Kay.⁷ Kay provides expressions for the functions $\sigma_+(\xi)$ and $\sigma_-(\xi)$.

Now, suppose we consider a general aperture distribution satisfying (28). This can be written as

$$f(x) = \int_{-\infty}^{\infty} A(\beta) \exp(-j\beta x + \alpha x) d\beta; \quad (34)$$

i.e., as a superposition of distributions of the form of (30). The field appropriate to this general aperture distribution will therefore be given by superposition of the solutions in (33); *i.e.*,

⁷ A. F. Kay, "Scattering of a surface wave by a discontinuity in reactance," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-7, pp. 22-31; January, 1959.

$$H_z = \frac{\omega\epsilon}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{A(\beta) \sigma_+(\beta + j\alpha) \exp[-j\xi x - j\sqrt{k^2 - \xi^2} y]}{(j\xi - j\beta + \alpha)(u - j\sqrt{k^2 - \xi^2})} d\xi d\beta. \quad (35)$$

This expression provides an exact solution to the launching problem which was posed. It is, however, not suited for numerical calculations because of the complexity of the function $\sigma_+(\beta + j\alpha)$. It might be thought that some simplification could be effected by evaluating the integration with respect to β by a contour method. Unfortunately, if β is regarded as a complex variable, $A(\beta)$ is regular in the lower half plane, whereas $\sigma_+(\beta + j\alpha)$ is regular in the upper half plane. This precludes any simple evaluation. Eq. (35) can, however, be taken as a starting point for the development of approximate solutions.

IV. APPENDIX

Justification of (1)

The result in (1) is suggested by the analysis given by Booker and Clemmow² and the integral term is identical to their result. It might be thought that this integral gives the total field, but it can be very easily shown that this is not so. Consider, for example, the situation when the aperture field corresponds to a pure surface wave; *i.e.*,

$$f(y) = \exp(-uy) \quad y \geq 0. \quad (36)$$

Then,

$$P(\xi) = \int_0^{\infty} \exp(j\xi y - uy) dy = \frac{1}{u - j\xi} \quad (37)$$

when

$$P(\xi) = \frac{\xi - ju}{\xi + ju} \quad P(-\xi) = \frac{1}{u + j\xi} + \frac{\xi - ju}{\xi + ju} \cdot \frac{1}{u + j\xi} = 0. \quad (38)$$

The integral, therefore, vanishes identically. It can be easily verified that the second term in (1) correctly represents the surface wave throughout the whole region considered.

Eq. (1) can be derived formally either by a suitable extension of the Booker-Clemmow method or by a manipulation of the result obtained by Clemmow. The validity of the result may be confirmed by establishing that:

- 1) the expression for H_z satisfies the wave equation (this is obvious from the form of the expression);
- 2) the boundary condition $j\omega\epsilon E_x = -uH_z$ when $y=0$ is satisfied; and
- 3) $E_y = f(y)$ when $x=0$ and $y \geq 0$.

From (1), and Maxwell's equations, we have

$$E_x = -\frac{j}{\omega\epsilon} \frac{\partial H_z}{\partial y}$$

$$= -\frac{1}{2\pi} \int_{-\infty}^{\infty} \left[P(\zeta) + \frac{\zeta - ju}{\zeta + ju} P(-\zeta) \right] \exp[-j\zeta y - j\sqrt{k^2 - \zeta^2}x] \frac{\zeta d\zeta}{\sqrt{k^2 - \zeta^2}} + \frac{2ju^2}{\beta} P(ju) \exp(-uy - j\beta x). \quad (39)$$

So, when $y=0$,

$$j\omega\epsilon E_x + uH_z$$

$$= \frac{\omega\epsilon}{2\pi} \int_{-\infty}^{\infty} (-j\zeta + u) \left[P(\zeta) + \frac{\zeta - ju}{\zeta + ju} P(-\zeta) \right] \exp(-j\sqrt{k^2 - \zeta^2}x) \frac{d\zeta}{\sqrt{k^2 - \zeta^2}} + \left[-\frac{2\omega\epsilon u^2}{\beta} + \frac{2\omega\epsilon u^2}{\beta} \right] P(ju) \exp(-j\beta x)$$

$$= \frac{\omega\epsilon}{2\pi j} \int_{-\infty}^{\infty} [(\zeta + ju)P(\zeta) + (\zeta - ju)P(-\zeta)] \exp(-j\sqrt{k^2 - \zeta^2}x) \frac{d\zeta}{\sqrt{k^2 - \zeta^2}}. \quad (40)$$

Consider

$$I = \int_{-\infty}^{\infty} (\zeta + ju)P(\zeta) \exp(-j\sqrt{k^2 - \zeta^2}x) \frac{d\zeta}{\sqrt{k^2 - \zeta^2}}, \quad (41)$$

and replace ζ by $-\zeta$, giving

$$I = - \int_{-\infty}^{\infty} (\zeta - ju)P(-\zeta) \exp(-j\sqrt{k^2 - \zeta^2}x) \frac{d\zeta}{\sqrt{k^2 - \zeta^2}}. \quad (42)$$

If (42) is subtracted from (41), it is seen that the integral in (40) vanishes and that boundary condition 2 is satisfied.

Also,

$$E_y = \frac{j}{\omega\epsilon} \frac{\partial H_z}{\partial x}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[P(\zeta) + \frac{\zeta - ju}{\zeta + ju} P(-\zeta) \right] \exp(-j\zeta y - j\sqrt{k^2 - \zeta^2}x) d\zeta + 2uP(ju) \exp(-uy - j\beta x). \quad (43)$$

Hence, when $x=0$,

$$E_y(0, y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[P(\zeta) + \frac{\zeta - ju}{\zeta + ju} P(-\zeta) \right] \exp(-j\zeta y) d\zeta + 2uP(ju) \exp(-uy). \quad (44)$$

Now, by inverting (4), we have

$$f(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} P(\zeta) \exp(-j\zeta y) d\zeta, \quad (45)$$

so that

$$E_y(0, y) = f(y) + I_1 + 2uP(ju) \exp(-uy), \quad (46)$$

where

$$I_1 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\zeta - ju}{\zeta + ju} P(-\zeta) \exp(-j\zeta y) d\zeta. \quad (47)$$

Since $f(y)$ is zero for $y \leq 0$, $P(\zeta)$ is regular and of algebraic growth, as $|\zeta|$ tends to infinity if $\text{Im } \zeta > -\alpha$ where α is some positive quantity. Hence, $P(-\zeta)$ is regular

and of algebraic growth as $|\zeta|$ tends to infinity, if $\text{Im } \zeta < \alpha$. When $y \geq 0$, the integral in (47) can, therefore, be evaluated by closing the contour in the lower-half plane, giving

$$I_1 = -\frac{1}{2\pi} [2\pi j \times \text{residue at the pole } \zeta = -ju]$$

$$= -j [-2juP(ju)] = -2uP(ju). \quad (48)$$

Substitution in (46) leads to

$$E_y(0, y) = f(y), \quad (49)$$

as required by condition (3).

The expression quoted in (1) therefore satisfies all the conditions required.

The Surface-Wave Concept in Connection With Propagation Trajectories Associated With the Sommerfeld Problem

H. BREMMER†

Summary—Both the continuous-wave solution and the pulse solution of the Sommerfeld problem can be represented as multiple integrals which describe cooperating contributions propagated along continuous sets of trajectories. The latter consist of a number of rectilinear sections that connect the transmitter and receiver. The propagation velocity along each section is in accordance with the corresponding medium. The velocity along section lying in the earth's surface is that of the well-known surface wave. Transient phenomena at the receiver start after the arrival of a main pulse along the trajectory connecting the transmitter and the receiver in accordance with ordinary geometric optics, ignoring surface-wave effects. Pulses along trajectories containing a surface-wave section may arrive earlier; however, their joint contributions then cancel each other until the arrival of the main pulse.

I. INTRODUCTION

THE question of the reality of the surface wave usually concerns the existence of a special contribution that has the characteristics of such a wave. Such a contribution, if any, then constitutes a part of the complete rigorous solution of the problem. The most striking property of the surface wave is that of its phase velocity. The latter is given by c/n_{12} for n_{12} defined by

$$\frac{1}{n_{12}^2} = \frac{1}{n_1^2} + \frac{1}{n_2^2}, \quad (1)$$

where n_1 and n_2 represent the refractive indices of the half spaces above and inside the earth.

The role of the velocity c/n_{12} is particularly clear for the pulse solution Π_s ; the latter corresponds to a transmitter moment with a time dependence proportional to a delta function. The pulse solution is simplest when both half spaces are assumed as dielectrics so as to have real-valued refractive indices n_1 , n_2 and n_{12} ; this situation constitutes a fair approximation for most microwaves. The pulse solution then becomes an elementary function if, moreover, the transmitter and receiver are placed on the earth's surface. For this situation, an investigation by Van der Pol [1] considers the related case of a transmitter moment with time dependence proportional to Heaviside's unit function; the pulse solution defined here is readily obtained from Van der Pol's solution by a differentiation with respect to the time.

The results arrived at by Van der Pol show explicitly the significance of the surface-wave phase velocity c/n_{12} . In the present paper, we derive representations for both the continuous-wave and the pulse solution, which clearly show the effect of this phase velocity in

the case of arbitrary positions of the transmitter and the receiver. However, the final expressions cannot be reduced to elementary functions under these most general circumstances.

An originally perfectly sharp impulse, radiated by the transmitter, is broadened to a transient phenomenon at the receiver. This broadening can be ascribed to delay effects suffered by contributions that have traveled along different propagation paths on their way from the transmitter to the receiver. A section of these paths lies in the earth's surface, but the effect of the corresponding surface-wave velocity c/n_{12} is masked by the interference due to the addition of the contributions of all paths. Further, the dispersion connected with the frequency dependence of n_2 will lead to another broadening if the finite conductivity of the earth has to be taken into account. Its effect, as well as that of any absorption in the atmosphere, is neglected throughout in the present article. We start with the special case of zero heights of transmitter and receiver.

II. A SURFACE-INTEGRAL REPRESENTATION OF THE CONTINUOUS-WAVE SOLUTION, BOTH TRANSMITTER AND RECEIVER ON THE GROUND

We assume a short vertical electric dipole at the origin (0,0,0), with a moment $Me^{-i\omega t}$. According to the Sommerfeld theory, the propagation across the earth's surface ($z=+0$) then depends on the following expression for the amplitude Π of the vertically directed Hertzian vector:

$$\Pi(x, y, +0; \omega) = \frac{2M}{c^3} \omega^2 \frac{k_2^2}{k_1^2} \int_0^\infty \frac{\lambda \cdot J_0(\lambda \sqrt{x^2 + y^2})}{k_2^2 \sqrt{\lambda^2 - k_1^2} + k_1^2 \sqrt{\lambda^2 - k_2^2}} d\lambda. \quad (2)$$

The wave numbers $k_1 = n_1(\omega/c)$ and $k_2 = n_2(\omega/c)$ refer to the atmosphere ($z > 0$) and the space inside the earth ($z < 0$), respectively; both half spaces are assumed as homogeneous with a plane interface. In either half space, the electric and magnetic fields are represented by

$$E = \frac{ic}{\omega} \text{curl curl } \Pi \quad \text{and} \quad H = n^2 \text{curl } \Pi$$

when using Gaussian units. In what follows, integrations through poles or branchpoints are avoided by assuming, if necessary, infinitesimal positive imaginary parts of the refractive indices and the wave numbers.

In order to transform (2) into an expression which is convenient for our purpose, we consider the two-dimensional image of (1) in the sense of operational calculus.

† Philips Res. Labs., N. V. Philips' Gloeilampenfabrieken, Eindhoven, Netherlands.

The latter connects any function $h(x, y)$ with an operational "image" defined by

$$f(p_1, p_2) = p_1 p_2 \iint_{-\infty}^{\infty} dx dy e^{-p_1 x - p_2 y} h(x, y),$$

which relation may be abbreviated as $h(x, y) \doteq f(p_1, p_2)$. The following general rule for radially symmetric functions is then easily verified for $\text{Re } p_1 = \text{Re } p_2 = 0$ (though the domain of applicability can be much larger):

$$\int_0^{\infty} \lambda J_0(\lambda \sqrt{x^2 + y^2}) f(\lambda) d\lambda \doteq 2\pi p_1 p_2 f(\sqrt{-p_1^2 - p_2^2}); \quad (3)$$

$f(\lambda)$ is assumed as an even function.

An application of (3) to (2) results in a relation which can be reduced to the following form:

$$\Pi(x, y, +0; \omega) \doteq 4\pi i \frac{M}{c} \frac{h^2}{n_1^2 k_1^2 (k_2^2 - k_1^2)} p_1 p_2 \frac{k_2^2 \sqrt{p_1^2 + p_2^2 + k_1^2} - k_1^2 \sqrt{p_1^2 + p_2^2 + k_2^2}}{p_1^2 + p_2^2 + h^2}, \quad (4)$$

in which $h = (\omega/c)n_{12}$ represents the wave-number of the surface wave; the imaginary parts of the square roots should be positive or zero.

The right-hand side of (4) can be split in many ways into factors which enable an application of the operational rule for composition (or convolution) products. In our case dealing with two variables, this rule reads [2]:

$$h_1(x, y) * h_2(x, y) = \iint_{-\infty}^{\infty} d\xi d\eta h_1(\xi, \eta) h_2(x - \xi, y - \eta) \doteq \frac{f_1(p_1, p_2) f_2(p_1, p_2)}{p_1 p_2}, \quad (5)$$

if the component factors satisfy the operational relations $h_1 \doteq f_1$ and $h_2 \doteq f_2$.

One out of the many possible factorizations of the $p_1 p_2$ function in (4) leads to the following representation replacing (2):

$$\Pi(x, y, +0; \omega) = \frac{2M}{c n_1^2} \frac{h^2}{k_1^2 (k_2^2 - k_1^2)} \left\{ \frac{k_2^2 e^{ik_1 \rho} - k_1^2 e^{ik_2 \rho}}{\rho} - \frac{i}{4} h^2 H_0^{(1)}(h\rho) * \frac{k_1^2 e^{ik_1 \rho} - k_2^2 e^{ik_2 \rho}}{\rho} \right\}, \quad (6)$$

with $\rho = (x^2 + y^2)^{1/2}$. The correctness of (6) can be verified by determining its operational image with the aid of (5), while applying, moreover, the two following important relations:

$$H_0^{(1)}(h\sqrt{x^2 + y^2}) \doteq 4i \frac{p_1 p_2}{p_1^2 + p_2^2 + h^2}, \quad (7)$$

$$\frac{\exp(ik\sqrt{x^2 + y^2 + z^2})}{\sqrt{x^2 + y^2 + z^2}}$$

$$\doteq 2\pi i p_1 p_2 \frac{\exp(i|z|\sqrt{p_1^2 + p_2^2 + k^2})}{\sqrt{p_1^2 + p_2^2 + k^2}}. \quad (8)$$

In its turn, these latter relations can be checked with the aid of (3) for $\text{Re } p_1 = \text{Re } p_2 = 0$, remembering the assumption of infinitesimal positive imaginary parts of the wave numbers. The final "image" of (6) then proves to be identical with the right-hand side of (4).

In view of (5), the composition product in (6) can be written out as a two-dimensional integral over the earth's surface, with surface element $dO_Q = d\xi d\eta$. The integrand then depends on the mutual distances TQ and QP between the transmitter at $T(0, 0, 0)$, the variable integration point at $Q(\xi, \eta, 0)$, and the receiver at $P(x, y, 0)$. Moreover, we pass from the wave numbers $k_{1,2} = n_{1,2}\omega/c$ and $h = n_{12}\omega/c$ to the refractive indices n_1, n_2 , and n_{12} . We thus obtain:

$$\begin{aligned} \Pi(x, y, +0; \omega) &= \frac{2M}{c} \frac{n_{12}^2}{n_1^4 (n_2^2 - n_1^2)} \left\{ \frac{n_2^2 e^{i(\omega/c)n_1 TP} - n_1^2 e^{i(\omega/c)n_2 TP}}{TP} \right. \\ &\quad \left. - \frac{i}{4} \frac{\omega^2}{c^2} n_{12}^2 \iint dO_Q H_0^{(1)} \left(\frac{\omega}{c} n_{12} TQ \right) \frac{n_1^2 e^{i(\omega/c)n_1 QP} - n_2^2 e^{i(\omega/c)n_2 QP}}{QP} \right\}. \quad (9) \end{aligned}$$

III. A SURFACE-INTEGRAL REPRESENTATION OF THE PULSE SOLUTION, BOTH TRANSMITTER AND RECEIVER ON THE GROUND

We next consider a dipole moment $M\delta(t)$ instead of $Me^{-i\omega t}$. The frequency spectrum of the delta function is given by

$$\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} d\omega.$$

Hence, the pulse solution Π_s is connected as follows with the continuous-wave solution (9):

$$\Pi_s(x, y, +0; t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Pi(x, y, +0; \omega) e^{-i\omega t} d\omega. \quad (10)$$

We substitute $\omega = iq$. The right-hand side of (10) then transforms into the inversion integral of the following one-dimensional operational relation connecting q with the time variable t :

$$\Pi_s(x, y, +0; t) \doteq q \Pi(x, y, +0; iq). \quad (11)$$

We substitute (9) in the right-hand side of (11), using the identity $H_0^{(1)}(iz) = (2/\pi i) K_0(z)$. The "original" with respect to q of the resulting expression can then be obtained with the aid of the operational relations:

$$q \doteq \delta(t),$$

$$qK_0\left(\frac{q}{c}n_{12}TQ\right) \doteq \frac{U\left(t - \frac{n_{12}}{c}TQ\right)}{\sqrt{t^2 - \frac{n_{12}^2}{c^2}TQ^2}}; \quad (12)$$

$U(x)$ denotes Heaviside's unit function (zero for $x < 0$, unity for $x > 0$). We also need the "shift rule" [according to which a factor $\exp(-\lambda q)$ corresponds, for the two-sided operational calculus used here, to replacing t by $t - \lambda$], and the differentiation rule (multiplication by q corresponding to the operator d/dt). These operations lead to the following final result, remembering the independence of the refractive indices $n_{1,2}$ and n_{12} of ω (in view of our assumption of a negligible conductivity of the earth):

$$\begin{aligned} \Pi_\delta(x, y, +0; t) = & \frac{2M}{cn_1^4} \frac{n_{12}^2}{(n_2^2 - n_1^2)} \left[\frac{n_2^2 \delta\left(t - \frac{n_1}{c}TP\right) - n_1^2 \delta\left(t - \frac{n_2}{c}TP\right)}{TP} \right. \\ & \left. + \frac{n_{12}^2}{2\pi c^2} \frac{d^2}{dt^2} \iint \frac{dO_Q}{QP} \left\{ \frac{n_1^2 U\left(t - \frac{n_{12}TQ + n_1QP}{c}\right)}{\sqrt{\left(t - \frac{n_1}{c}QP\right)^2 - \frac{n_{12}^2}{c^2}TQ^2}} - \frac{n_2^2 U\left(t - \frac{n_{12}TQ + n_2QP}{c}\right)}{\sqrt{\left(t - \frac{n_2}{c}QP\right)^2 - \frac{n_{12}^2}{c^2}TQ^2}} \right\} \right] \quad (13) \end{aligned}$$

Obviously, the first term represents an undistorted pulse, propagating from T to P with the phase velocity c/n_1 holding for the upper medium; the second term corresponds to a similar pulse with the phase velocity c/n_2 of the lower medium. In other words, these terms can be interpreted as disturbances traveling just above and just below the earth's surface. On the other hand, the argument of the unit-function in the third term in-

with the velocity c/n_1 of the upper medium over the section QP . Therefore, the third term of (13) represents contributions traveling as a surface wave up to some point Q on the earth's surface, and thence as an ordinary wave through the upper medium, directly above the earth, towards P . Similarly, the fourth term of (13) represents a contribution for which the second section QP of the propagation path lies in the space directly underneath the earth's surface (corresponding to the phase velocity c/n_2).

This interpretation of the four contributions to (13) is represented schematically in Fig. 1, which also indi-

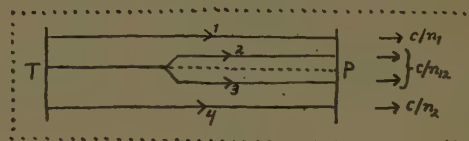


Fig. 1.

cates the corresponding phase velocities. We observe that the succession of the two sections in the third and fourth terms may just as well be inverted.

IV. THE REDUCTION OF THE PREVIOUS PULSE SOLUTION TO ELEMENTARY FUNCTIONS

The two surface integrals of (13) equal the two following elementary functions:

$$n_1^2 \iint \frac{dO_Q}{QP} \frac{U\left(t - \frac{n_{12}TQ + n_1QP}{c}\right)}{\sqrt{\left(t - \frac{n_1}{c}QP\right)^2 - \frac{n_{12}^2}{c^2}TQ^2}} = 2\pi c \frac{n_1 n_2}{n_{12}} U\left(t - \frac{n_{12}}{c}TP\right) \operatorname{arcosh} \min\left(\frac{ct}{n_{12}TP}, \frac{n_1}{n_{12}}\right), \quad (14)$$

$$n_2^2 \iint \frac{dO_Q}{QP} \frac{U\left(t - \frac{n_{12}TQ + n_2QP}{c}\right)}{\sqrt{\left(t - \frac{n_2}{c}QP\right)^2 - \frac{n_{12}^2}{c^2}TQ^2}} = 2\pi c \frac{n_1 n_2}{n_{12}} U\left(t - \frac{n_{12}}{c}TP\right) \operatorname{arcosh} \min\left(\frac{ct}{n_{12}TP}, \frac{n_2}{n_{12}}\right). \quad (15)$$

indicates a propagation with the surface-wave velocity c/n_{12} over the distance TQ , followed by a propagation

These identities hold in general, provided that both T and P are situated in the plane of integration $z=0$.

We shall briefly discuss a derivation of (15), that of (14) being completely analogous in view of the symmetry with respect to n_1 and n_2 . The right-hand side of (15) can be put into the form:

$$\frac{2\pi c n_1 n_2}{n_{12}} U\left(t - \frac{n_{12}}{c} TP\right) \left\{ \int_1^{n_2/n_{12}} \frac{ds}{\sqrt{s^2 - 1}} + U\left(\frac{n_2}{c} TP - t\right) \int_{n_2/n_{12}}^{ct/(n_{12}TP)} \frac{ds}{\sqrt{s^2 - 1}} \right\},$$

$$\iint \frac{dO_Q}{QP} \left\{ \frac{n_1^2 U\left(t - \frac{n_{12}TQ + n_1QP}{c}\right)}{\sqrt{\left(t - \frac{n_1}{c} QP\right)^2 - \frac{n_{12}^2}{c^2} TQ^2}} - \frac{n_2^2 U\left(t - \frac{n_{12}TQ + n_2QP}{c}\right)}{\sqrt{\left(t - \frac{n_2}{c} QP\right)^2 - \frac{n_{12}^2}{c^2} TQ^2}} \right\} \\ = \begin{cases} 0 & \text{for } t < \frac{n_1}{c} TP, \\ 2\pi c \frac{n_1 n_2}{n_{12}} \left\{ \text{arc cosh}\left(\frac{n_1}{n_{12}}\right) - \text{arc cosh}\left(\frac{ct}{n_{12}TP}\right) \right\} & \text{for } \frac{n_1}{c} TP < t < \frac{n_2}{c} TP, \\ 2\pi c \frac{n_1 n_2}{n_{12}} \left\{ \text{arc cosh}\left(\frac{n_2}{n_{12}}\right) - \text{arc cosh}\left(\frac{n_2}{n_{12}}\right) \right\} & \text{for } t > \frac{n_2}{c} TP. \end{cases} \quad (16)$$

which admits an easy determination of its q image (with respect to t); for the second term, this can be achieved by inverting the order of integration in the corresponding definition integral of the operational calculus. The operational image of the left-hand side of (15) can also be determined, with the aid of (12) and the shift rule. This identity (15) thus proves to be equivalent to the following one in terms of the variable q

$$q \iint \frac{dO_Q}{QP} \cdot K_0\left(\frac{q}{c} n_{12} TQ\right) e^{-(n_2/c)qQP} \\ = \frac{2\pi c}{n_2 n_{12}} \int_1^{n_2/n_{12}} \frac{e^{-qn_{12}(TP/c)s}}{\sqrt{s^2 - 1}} ds.$$

The left-hand side here constitutes a composition product with respect to the coordinates x and y of P if T is still assumed at the origin. We next return from q to $\omega = iq$, and also from the refractive indices to the wave numbers $k_{1,2} = (\omega/c)n_{1,2}$ and $h = n_{12}(\omega/c)$. We then arrive at another identity which can be formulated as follows:

$$H_0^{(1)}(h\sqrt{x^2 + y^2}) * \frac{\exp(ik_2\sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}} \\ = \frac{4k_1}{ih^2 k_2} \int_{s=1}^{s=k_2/h} \frac{1}{\sqrt{s^2 - 1}} d \left\{ \frac{\exp ihs\sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2}} \right\}.$$

Finally, this relation can be transposed, with the aid of (5), (7) and (8), into its two-dimensional "image" with respect to x and y . This leads to an elementary identity in the quantity $p_1^2 + p_2^2$, which ultimately proves the correctness of (15).

The integral in (13) now depends on the difference of (14) and (15). We assume $n_2 > n_1$; in view of the other inequalities $n_{12} < n_{1,2}$ [see (1)], this difference can be reduced to the further identity:

This relation admits a remarkable interpretation. Obviously, the effects of the two individual integrals [corresponding to the third and fourth terms of (13)] start at the moments

$$t_3 = \frac{1}{c} \min (n_{12}TQ + n_1QP)$$

and

$$t_4 = \frac{1}{c} \min (n_{12}TQ + n_2QP);$$

min is meant here with respect to a variation of Q across the plane $z=0$. When taking into account the inequalities $n_{12} < n_{1,2}$, both instants prove to be identical with the time $(1/c)n_{12}TP$. Nevertheless, the combined effect represented by (16) only starts at the later time $(n_1/c)TP$. Hence, the joint effect of the third and fourth terms of (13) (see Fig. 1) is such that no contribution whatever remains between the times of arrival of the surface wave and the wave propagating through the upper medium.

We can now evaluate the second-order time derivative in (13) with the aid of (16), when also applying the properties

$$\frac{dU(x)}{dx} = \delta(x) \quad \text{and} \quad h(x)\delta(x) = h(0)\delta(x).$$

This leads to the final representation

$$\begin{aligned} \Pi_3(x, y, +0; t) \\ = -\frac{2M}{c^2} \frac{n_2 n_1^3}{n_1^3(n_2^2 - n_1^2)} \left\{ U\left(t - \frac{n_1}{c} TP\right) \right. \\ \left. - U\left(t - \frac{n_2}{c} TP\right) \right\} \cdot \frac{d}{dt} \frac{1}{\sqrt{t^2 - \frac{n_1^2}{c^2} TP^2}}. \quad (17) \end{aligned}$$

Therefore, the response of a transmitter moment proportional to a delta function is restricted, on the earth's surface, to the finite time interval between the arrivals of the waves propagated through the upper medium (velocity c/n_1) and through the lower medium (velocity

$$\begin{aligned} \Pi(x, y, z; \omega) = \frac{i}{4\pi n_1^2} \frac{M}{c} \frac{h^2}{(k_2^2 - k_1^2)} \left(\frac{\partial^3}{\partial h_1^2 \partial z} + \frac{k_1^2}{k_2^2} \frac{\partial^3}{\partial h_1 \partial z^2} \right) \\ \cdot \left\{ \frac{\exp(i k_1 \sqrt{x^2 + y^2 + h_1^2})}{\sqrt{x^2 + y^2 + h_1^2}} * H_0^{(1)}(h \sqrt{x^2 + y^2}) * \frac{\exp(i k_2 \sqrt{x^2 + y^2 + z^2})}{\sqrt{x^2 + y^2 + z^2}} \right\}. \quad (20) \end{aligned}$$

c/n_2). The solution (17) constitutes the time derivative of the response function investigated by Van der Pol [3].

V. THE SOLUTION INSIDE THE EARTH REPRESENTED AS A JOINT CONTRIBUTION FROM VARIOUS PROPAGATION PATHS

The method outlined in Sections II and III can be repeated for T at an arbitrary height h_1 above the earth's surface, and P above or below it. In the latter case, the Sommerfeld theory involves an expression which can be represented as follows in terms of its operational image with respect to the coordinates x and y of the projection $P'(x, y, 0)$, on the earth's surface, of the receiver at $P(x, y, z)$:

$$\begin{aligned} \Pi(x, y, z; \omega) \doteq 4\pi i \frac{M}{c n_1^2} \frac{h^2 p_1 p_2}{k_2^2 (k_2^2 - k_1^2)} \frac{k_2^2 \sqrt{p_1^2 + p_2^2 + k_1^2} - k_1^2 \sqrt{p_1^2 + p_2^2 + k_2^2}}{p_1^2 + p_2^2 + h^2} \\ \cdot \exp \{ i(h_1 \sqrt{p_1^2 + p_2^2 + k_1^2} + |z| \sqrt{p_1^2 + p_2^2 + k_2^2}) \} \quad (18) \end{aligned}$$

This expression can also be made plausible as follows. In the special case $h_1 = 0$ it follows from (4) after a multiplication by $(k_1^2/k_2^2) \exp(i|z| \sqrt{p_1^2 + p_2^2 + k_2^2})$. The first factor k_1^2/k_2^2 originates from the boundary condition requiring

$$k_1^2 \Pi(x, y, +0; \omega) = k_2^2 \Pi(x, y, -0; \omega);$$

the further exponential factor expresses that the field for $z < 0$ is obtained from that for $z = -0$ after a multiplication of the p image by $\exp(-iz \sqrt{p_1^2 + p_2^2 + k_2^2})$. The latter factor amounts to a multiplication of Π itself by the operator

$$\exp - \{ iz \sqrt{\partial^2/\partial x^2 + \partial^2/\partial y^2 + k_2^2} \}, \quad (19)$$

the significance of which is apparent for any part of the two-dimensional Fourier synthesis of $\Pi(x, y, -0)$. The

operator in question guarantees that $\Pi(x, y, z)$, derived from (18), satisfies both the wave equation for $z < 0$, and the radiation condition at $z \rightarrow -\infty$. The remaining factor $\exp(ih_1 \sqrt{p_1^2 + p_2^2 + k_1^2})$ accounts for the influence of the finite transmitter height. In view of the reciprocity theorem, this influence will be the same as that of raising the receiver in the upper medium from $z = +0$ to $z = h_1$; in analogy to (19), the latter raising (taking place in the medium with wave number k_1) would result in the factor just mentioned.

One of the possible factorizations of the right-hand side of (18) involves the following composition product of three factors:

This relation can be checked by the same method described for the similar relation (6). We then have to apply, instead of (5), its extension for three factors, *viz.*,

$$\begin{aligned} h_1(x, y) * h_2(x, y) * h_3(x, y) \\ = \int \int_{-\infty}^{\infty} d\xi_1 d\eta_1 \cdot h_1(\xi_1, \eta_1) \int \int_{-\infty}^{\infty} d\xi_2 d\eta_2 \cdot h_2(\xi_2, \eta_2) \\ \cdot h_3(x - \xi_1 - \xi_2, y - \eta_1 - \eta_2) \\ \doteq \frac{f_1(p_1, p_2) f_2(p_1, p_2) f_3(p_1, p_2)}{p_1^2 p_2^2}. \quad (21) \end{aligned}$$

The differentiations with respect to h_1 and z yield no difficulties when verifying (20); in fact, these variables

are to be considered as parameters independent of the operational transformations.

We can next pass from (20) to a representation in terms of a surface integral connected with two successive integrations over the earth's surface $z = 0$. The identity to be applied, *viz.*,

$$\begin{aligned} h_1(\sqrt{x^2 + y^2}) * h_2(\sqrt{x^2 + y^2}) * h_3(\sqrt{x^2 + y^2}) \\ = \int \int_{z_{Q_1}=0} dO_1 \int \int_{z_{Q_2}=0} dO_2 h_1(OQ_1) h_2(Q_1 Q_2) h_3(Q_2 P_0), \quad (22) \end{aligned}$$

holds in view of the radial symmetry of the relevant functions $h_i(x, y) = h_i(\sqrt{x^2 + y^2})$. In this relation dO_1 and dO_2 represent surface elements referring to the two independent integration points Q_1 and Q_2 , $Q_2 P_0$ the distance from Q_2 to $P_0(x, y, 0)$, and OQ_1 the distance

from the origin to Q_1 . The application of (22) leads to the following alternative expression for the continuous-wave solution (20), if we introduce once more the refractive indices $n_{1,2}$, n_1 and n_2

$$\begin{aligned} \Pi(P; \omega) &= \frac{i}{4\pi} \frac{M}{c} \frac{n_{12}^2}{(n_2^2 - n_1^2)} \left(\frac{1}{n_1^2} \frac{\partial^3}{\partial h_1^2 \partial z_P} + \frac{1}{n_2^2} \frac{\partial^3}{\partial h_1 \partial z_P^2} \right) \\ &\quad \left\{ \iint dO_1 \iint dO_2 \frac{e^{i(\omega/c)n_1 \cdot TQ_1}}{TQ_1} \cdot H_0^{(1)} \left(\frac{\omega}{c} n_{12} Q_1 Q_2 \right) \cdot \frac{e^{i(\omega/c)n_2 \cdot Q_2 P}}{Q_2 P} \right\}; \quad (23) \end{aligned}$$

$T(0, 0, h_1)$ and $P(x, y, z)$ represent the transmitter and the receiver, respectively.

The corresponding pulse solution can be evaluated according to (11), along the same lines which led from (9) to (13). The final result reads:

$$\begin{aligned} \Pi_s(P; t) &= \frac{M}{2\pi^2 c} \frac{n_{12}^2}{(n_2^2 - n_1^2)} \left(\frac{1}{n_1^2} \frac{\partial^3}{\partial h_1^2 \partial z_P} + \frac{1}{n_2^2} \frac{\partial^3}{\partial h_1 \partial z_P^2} \right) \\ &\quad \cdot \iint \frac{dO_1}{TQ_1} \iint \frac{dO_2}{Q_2 P} \frac{U \left(t - \frac{n_1 TQ_1 + n_{12} Q_1 Q_2 + n_2 Q_2 P}{c} \right)}{\sqrt{\left(t - \frac{n_1 TQ_1 + n_2 Q_2 P}{c} \right)^2 - \frac{n_{12}^2}{c^2} Q_1 Q_2^2}}. \quad (24) \end{aligned}$$

Here we recognize the superposition of contributions propagated along trajectories consisting of the following successive parts: 1) the section TP through the upper medium, from the transmitter towards an arbitrary point Q_1 on the earth's surface (propagation velocity c/n_1), 2) the section $Q_1 Q_2$ thence to another point Q_2 on the earth's surface (surface-wave propagation velocity c/n_{12}), 3) the section $Q_2 P$ through the lower medium towards the receiver (propagation velocity c/n_2).

VI. A SIMPLIFIED REPRESENTATION OF THE SOLUTION INSIDE THE EARTH

The reduction of the representation (13) of the pulse solution (transmitter and receiver on the earth) to the simpler form (17) showed in particular that the starting time of the transient phenomenon is given by the arrival of the wave traveling through the upper medium, and not by the earlier arrival of the surface wave. Similar circumstances occur in the half space $z < 0$. The expres-

sion (24) would suggest a starting of the transient at the time

$$\frac{1}{c} \min (n_1 TQ_1 + n_{12} Q_1 Q_2 + n_2 Q_2 P), \quad (25)$$

but an alternative representation discussed hereafter demonstrates that the starting time may be postponed, due to a complete canceling of the various contributions contained in (24).

The representation in question can be derived from the identity

$$\begin{aligned} \Pi(P; \omega) &= - \frac{n_1^2}{4\pi^2 n_2^2} \frac{\partial^2}{\partial h_1 \partial z_P} \iint dO_1 \iint dO_2 \frac{e^{i(\omega/c)n_1 TQ_1}}{TQ_1} \\ &\quad \cdot \Pi_{z=h_1 \rightarrow 0}(Q_1 Q_2; \omega) \frac{e^{i(\omega/c)n_2 Q_2 P}}{Q_2 P}, \quad (26) \end{aligned}$$

in which the Π factor in the integrand denotes the solution on the earth's surface (see Section II) for TP re-

placed by $Q_1 Q_2$. The relation (26) is verified by considering the integral as a composition product of three factors, the $p_1 p_2$ image of which can be determined with the aid of (8), (4) and the composition-product rule (21); the image then proves to be identical with (18).

The substitution $\omega = iq$ in (26) yields, after a multiplication by q , the q image of the corresponding pulse solution [see (11)]. It involves the following further relation, obtained with the aid of the shift rule of the operational calculus:

$$\begin{aligned} \Pi_s(P; t) &= - \frac{n_1^2}{4\pi^2 n_2^2} \frac{\partial^2}{\partial h_1 \partial z_P} \iint \frac{dO_1}{TQ_1} \iint \frac{dO_2}{Q_2 P} \\ &\quad \cdot \Pi_{z=h_1 \rightarrow 0, \delta}(Q_1 Q_2, t - \frac{n_1 TQ_1 + n_2 Q_2 P}{c}). \end{aligned}$$

We next substitute (17) and then arrive at the following final expression in which the effect of the unit functions is indicated by an inequality fixing the restricted domain of integrations:

$$\begin{aligned} \Pi_s(P; t) &= \frac{M}{2\pi^2 c^2} \frac{n_{12}^3}{n_1 n_2 (n_2^2 - n_1^2)} \frac{\partial^2}{\partial h_1 \partial z_P} \iint \frac{dO_1}{TQ_1} \iint \frac{dO_2}{Q_2 P} \frac{d}{dt} \frac{1}{\sqrt{\left(t - \frac{n_1 TQ_1 + n_2 Q_2 P}{c} \right)^2 - \frac{n_{12}^2}{c^2} Q_1 Q_2^2}} \\ &\quad \cdot \underbrace{\left(n_1 (TQ_1 + Q_1 Q_2) + n_2 Q_2 P < ct < n_1 TQ_1 + n_2 (Q_1 Q_2 + Q_2 P) \right)}_{\text{domain of integration}}. \quad (27) \end{aligned}$$

The domain of integration of this new representation indicates the actual starting time, *viz.*,

$$\frac{1}{c} \min (n_1 T Q_1 + n_1 Q_1 Q_2 + n_2 Q_2 P) = \frac{1}{c} \min (n_1 T Q_2 + n_2 Q_2 P), \quad (28)$$

VII. THE PULSE SOLUTION IN THE HALF SPACE ABOVE THE EARTH

The investigation of the continuous-wave solution in this space ($Z_p > 0$) can be based on the following $p_1 p_2$ image for the difference of the total field and the primary field (M/c) $\exp (ik_0 TP)/TP$:

$$\Pi_{\text{sec}}(x, y, z; \omega) \doteq 2\pi i \frac{M}{cn_1^2} \frac{h^2 p_1 p_2}{k_1^2 k_2^2 (k_2^2 - k_1^2)} \frac{(k_2^2 \sqrt{p_1^2 + p_2^2 + k_1^2} - k_1^2 \sqrt{p_1^2 + p_2^2 + k_2^2})^2}{(p_1^2 + p_2^2 + h^2) \sqrt{p_1^2 + p_2^2 + k_1^2}} \exp \{i(z+h_1) \sqrt{p_1^2 + p_2^2 + k_1^2}\} \quad (29)$$

instead of (25). This latter time points to a first pulse arriving along the Snell trajectory connecting T and P . In fact, if (28) constitutes a later time than (25), all contributions of (24) arriving between the moments (25) and (28) cancel one another. This situation occurs in the region outside the cone formed by those rays that originate after refraction from the rays leaving the transmitter under the Brewster angle.

This relation represents, in a condensed form similar to (18), the complete Sommerfeld theory for the upper half space; it can be made plausible by applying an operator analogous to (19) (replacing z by $-z$ and k_2 by k_1) to the secondary field on the earth. With the aid of (29) we can again derive integral representations for both the continuous-wave solution and the pulse solution. We are interested in two particular representations for the latter that correspond to (24) and (27). Since these representations can be verified along the same lines as discussed in the previous sections, we only give the final formulas, *viz.*,

$$\begin{aligned} \Pi_{\text{sec}, \delta}(P; t) = & -\frac{1}{4\pi^3} \frac{M}{c} \frac{n_{12}^2}{n_1^2 (n_2^2 - n_1^2)} \\ & \times \left[\pi \left\{ \left(\frac{n_1^2}{n_2^2} + \frac{n_2^2}{n_1^2} \right) \frac{\partial^3}{\partial h_1^3} + \frac{n_1^2 (n_2^2 - n_1^2)}{n_2^2 c^2} \frac{\partial^3}{\partial t^2 \partial h_1} \right\} \iint \frac{dO_1}{TQ_1} \iint \frac{dO_2}{Q_2 P} \frac{U \left(t - \frac{n_1 T Q_1 + n_{12} Q_1 Q_2 + n_1 Q_2 P}{c} \right)}{\sqrt{\left(t - n_1 \frac{T Q_1 + Q_2 P}{c} \right)^2 - \frac{n_{12}^2}{c^2} Q_1 Q_2^2}} \right. \\ & \left. + \frac{\partial^2}{\partial h_1^2} \left\{ \frac{\partial^2}{\partial h_1^2} + \frac{(n_2^2 - n_1^2)}{c^2} \frac{\partial^2}{\partial t^2} \right\} \iint \frac{dO_1}{TQ_1} \iint \frac{dO_2}{Q_1 Q_2} \iint \frac{dO_3}{Q_3 P} \frac{U \left(t - \frac{n_1 T Q_1 + n_2 Q_1 Q_2 + n_{12} Q_2 Q_3 + n_1 Q_3 P}{c} \right)}{\sqrt{\left(t - \frac{n_1 T Q_1 + n_2 Q_1 Q_2 + n_1 Q_3 P}{c} \right)^2 - \frac{n_{12}^2}{c^2} Q_2 Q_3^2}} \right]. \quad (30) \end{aligned}$$

$$\begin{aligned} \Pi_{\delta}(P; t) = & \frac{M}{cn_1^2} \left\{ \frac{\delta \left(t - \frac{n_1}{c} TP \right)}{TP} - \frac{\delta \left(t - \frac{n_1}{c} T' P'' \right)}{T' P''} \right\} \\ & + \frac{M}{\pi c^2} \frac{n_2 n_{12}^3}{n_1^3 (n_2^2 - n_1^2)} \frac{\partial}{\partial h_1} \iint \frac{dO_Q}{Q P''} \frac{d}{dt} \frac{1}{\sqrt{\left(t - \frac{n_1}{c} Q P'' \right)^2 - \frac{n_{12}^2}{c^2} T' Q^2}}. \quad (31) \end{aligned}$$

$n_1 T' Q + n_1 Q P'' < ct < n_2 T' Q + n_1 Q P''$

In the latter expression, T' denotes the projection of T on the earth's surface (the origin in the previous computations), and P'' the point $(x, y, z+h_1)$ which is on the same vertical as the receiver, but at a height h_1 above it. The fact that P is now assumed in the same half space as the transmitter involves the possibility of a representation which depends only on a single two-dimensional integration over the earth's surface.

According to (30), the secondary field can be considered as composed of two contributions which can be characterized as follows (see Fig. 2):

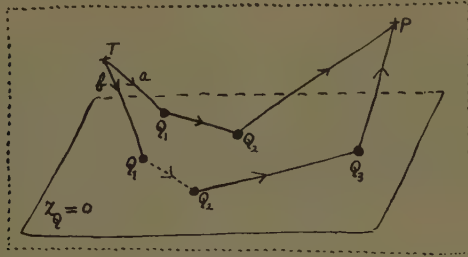


Fig. 2.

1) a first contribution with propagation path TQ_1Q_2P via two arbitrary points Q_1 and Q_2 on the earth's surface. The section Q_1Q_2 refers to a surface-wave propagation;

2) a second contribution with propagation path $TQ_1Q_2Q_3P$ via three consecutive points Q_1 , Q_2 and Q_3 on the earth's surface. The propagation velocities c/n_2 and c/n_{12} along Q_1Q_2 and Q_2Q_3 , respectively, indicate a propagation through the lower medium (just below the earth's surface) for the first section, and a surface-wave propagation for the other section. The order of these two sections is irrelevant.

The other expression (31) comprises two undistorted pulses and a transient phenomenon. The former originate from T and T' respectively. The pulse from T' to P'' can be identified with a propagation along the trajectory connecting T and P with an intermediate reflection against the earth's surface.

The transient described by the integral of (31) starts at the time

$$\frac{n_1}{c} \min (T'Q + QP'') = \frac{n_1}{c} T'P'', \quad (32)$$

that is, immediately after the passage at P of the second pulse. A comparison with (30) proves the mutual

cancellation by interference of all those contributions to (30) that arrive between the moments

$$\frac{1}{c} \min (n_1TQ_1 + n_2Q_1Q_2 + n_{12}Q_2Q_3 + n_1Q_3P) \\ = \frac{1}{c} \min (n_1TQ_1 + n_{12}Q_1Q_3 + n_1Q_3P), \quad (33)$$

and (32), provided that (33) is earlier than (32). This occurs in the region outside the cone formed by those reflected rays that originate from primary rays leaving the transmitter under the Brewster angle (compare with the end of the preceding section).

VIII. FINAL REMARKS

Expressions derived in this article indicate the existence of individual field contributions that are propagated along continuous sets of trajectories from the transmitter T to the receiver P . The pulse solutions can be represented in forms [see (13), (24) and (30)] which show the associated delay times explicitly in the argument of a Heaviside's unit function. The trajectories may contain a section in the earth's surface; the propagation velocity along such a section is that of a surface-wave. The transient effects connected with the pulse solution always start after the arrival of a main pulse propagated along the geometric-optical trajectory connecting T and P after a refraction or reflection. Contributions associated with trajectories containing a surface-wave section may arrive earlier; however, their effects are then cancelled by mutual interference until the arrival of the main pulse. Such a cancellation occurs in the regions reached by the refracted and reflected rays that originate from primary rays leaving the transmitter in directions less steep than that of the Brewster angle (see the remarks at the ends of Sections VI and VII).

The outlined methods are also applicable to the theory of a vertical magnetic dipole. The well-known absence there of surface-wave effects is then confirmed. In fact, the propagation trajectories here never contain a surface-wave section; therefore, the phase velocity c/n_{12} of these sections occurs nowhere in the explicit formulas.

REFERENCES

- [1] Balh. Van der Pol, "On discontinuous electromagnetic waves and the occurrence of a surface wave," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-4, pp. 288-293; July, 1956.
- [2] Balh. Van der Pol and H. Bremmer, "Operational Calculus," Cambridge University Press, Cambridge, Eng. p. 338; 1955.
- [3] Balh. Van der Pol, *op. cit.*, formula (17).

The Transmission Characteristics of a Corrugated Guide

GERHARD PIEFKE†

Summary—The transmission characteristics of a corrugated guide are analyzed under the assumption that the guide wavelength is always much greater than the "corrugation constant" ($D_1 + D_2$ of Fig. 1). The corrugated guide is, therefore, replaced by a quasi-homogeneous but anisotropic medium whose dielectric constant and permeability are represented by tensors.

It turns out that the corrugated guide behaves much like the Goubau guide if the corrugation depth is small with respect to the wavelength. Corresponding to the influence of the dielectric on the Goubau guide, increasing depth of the corrugations increases also, on this guide, the attenuation of the waves, decreases the phase velocity, and increases the field concentration around the guide accordingly. As a rule, however, the corrugated guide acts as a low pass whose pass bands are about at $k \cdot 0.5 < d/\lambda_0 < (0.25 + k \cdot 0.5)$ and whose stop bands are about at $(0.25 + k \cdot 0.5) < d/\lambda_0 < 0.5(1 + k)$ ($k = 0, 1, 2, \dots$, d = corrugation depth, λ_0 = wavelength of a plane wave in space).

Since, in the case of long radio waves, parallel building fronts and mountain ranges may be thought of as corrugated guides, one of the results of the paper is the fact that the propagation of waves around the earth cannot only be considerably curbed by building fronts and mountain ranges, but can even be suppressed altogether.

LIST OF PRINCIPAL SYMBOLS

- ϵ_0 = dielectric constant of space.
- μ_0 = permeability of space.
- $Z_0 = \sqrt{\mu_0/\epsilon_0}$ = field impedance of space.
- ϵ_i = dielectric constant of the dielectric in the corrugations.
- κ = conductivity of the metal.
- f = frequency.
- $\omega = 2\pi f$ = angular frequency.
- $\beta_0 = 2\pi/\lambda_0 = \omega\sqrt{\mu_0\epsilon_0}$ phase constant of a plane wave in space.
- λ_0 = wavelength of a plane wave in space.
- $\gamma \equiv \pm j(\beta - j\alpha)$ = axial propagation constant of the waves on the guide.
- $\beta = 2\pi/\lambda$ phase constant.
- α = attenuation constant.
- λ = wavelength of the guide modes.
- v_p = phase velocity.
- c = velocity of light.
- a = inner radius of the guide.
- b = outer radius of the guide.
- d = depth of the corrugations.
- D_1 = width of the corrugations.
- D_2 = spacing of the corrugations.
- $D_1 + D_2$ = corrugation constant (analogue to optics).
- ϑ = equivalent thickness of the conducting layer of the metal.
- r, ϕ, z = cylindrical coordinates.
- $k_0 b$ = eigenvalue associated with the respective mode.
- J_m = Bessel function of m th order.

J_m' = derivative of the Bessel function with respect to the argument.

N_m = Neumann function of m th order.

$H_m^{(1)}$ = Hankel function of first kind and m th order.

$H_m^{(2)}$ = Hankel function of second kind and m th order.

$H_m^{(1)'}$, $H_m^{(2)'}$ = derivations, with respect to the argument, of the Hankel functions of first and second kind.

INTRODUCTION

A PARTICULARLY important problem of communications is the propagation of surface modes along single wires as well as along the surface of the earth. The transmission characteristics of a single wire have been investigated in particular by Sommerfeld,¹ Hondros,² Harms,³ and Goubau.⁴

The transmission characteristics of a plane and cylindrical corrugated guide (see Fig. 1) have been described

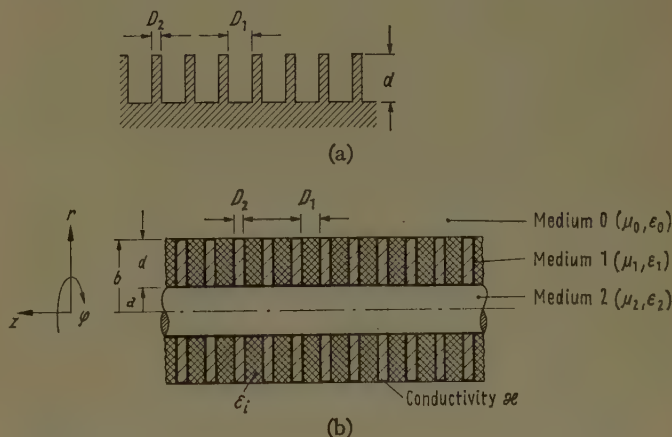


Fig. 1—(a) Plane; (b) cylindrical corrugated guides.

by Rotman,⁵ among others. The paper by Rotman⁵ indeed communicates formulas for calculation of the phase constants of the modes (e.g., as devised by Cutler), but does not present formulas for calculation of the attenuation constants. In the propagation of waves around the globe, parallel building fronts and mountain ranges may often be thought of as a plane corrugated

¹ A. Sommerfeld, "Vorlesungen über theoretische Physik, vol. III: Elektrodynamik," Dieterichsche Verlagsbuchhandlung, Wiesbaden, Germany, pp. 177–192; 1948.

² D. Hondros, "Über elektromagnetische Drahtwellen," *Annalen der Physik*, vol. 30, pp. 905–950; December, 1909.

³ F. Harms, "Elektromagnetische Wellen an einem Draht mit isolierender zylindrischer Hülle," *Ann. Physik*, vol. 23, pp. 44–60; May, 1907.

⁴ G. Goubau, "Surface waves and their application to transmission lines," *J. Appl. Phys.*, vol. 21, pp. 1119–1128; November, 1950.

⁵ W. Rotman, "A study of single-surface corrugated guides," *Proc. IRE*, vol. 39, pp. 952–959; August, 1951.

guide. Because of the high importance of the corrugated guide, the characteristics of this guide will be explained in detail in this paper, and formulas will be derived for calculation of the attenuation constants.

It is assumed that the guide wavelength λ always is much greater than $D_1 + D_2$ (see Fig. 1). According to Piefke⁶ the corrugated guide can then be thought of as a quasi-homogeneous but anisotropic medium in the range $a \leq r \leq b$ (medium 1 of Fig. 1). In medium 1 the dielectric constant and the permeability of the guide are then tensors; *i.e.*, excitation and field are no longer parallel vectors. The introduction of tensors offers the advantage that the wave equation can be easily solved also in medium 1 and the boundary conditions can be easily satisfied.

RESULTS OF THE ANALYSIS AND PHYSICAL EXPLANATIONS

The Tensors of the Dielectric Constant and the Permeability in the Anisotropic Medium 1

It is assumed that the inequalities

$$D_1 + D_2 \ll \lambda, \quad D_1 \ll \frac{1}{2}\lambda_0 \sqrt{\epsilon_0/\epsilon_i}, \quad D_2 \gg \vartheta \quad (1), (2), (3)$$

hold for the corrugated guide of Fig. 1. Because of the inequalities (1) and (2), medium 1 may be thought of as quasi-homogeneous but anisotropic.⁷ Dielectric constant and permeability are then tensors whose components are obtained from the following consideration:

Let us visualize in a Cartesian coordinate system an arrangement corresponding to medium 1 (see Fig. 2). It consists of metal plates which are arranged perpendicularly to the z -axis and which are insulated from each other by a dielectric. Corresponding to Fig. 1, the metal plates have the thickness D_2 , the dielectric has the thickness D_1 , and the separation $D_1 + D_2$ again is much smaller than the wavelength (1).

Now if a plane wave, with the electric field in *parallel* to the plates, strikes the medium in the direction y [see Fig. 2(a)] no wave can propagate between the metal plates because of (2). With high conductivity of the metal in parallel to the plates, the metal-plate medium of Fig. 2 is thus a conductor with the conductivity

$$\kappa_p = \kappa \frac{D_2}{D_1 + D_2} \quad (4)$$

or the dielectric constant

$$\epsilon_p = \kappa_p / j\omega. \quad (5)$$

⁶ G. Piefke, "Wellenausbreitung in der Scheiben-Leitung," *Arch. elektr. Übertrag.*, vol. 11, pp. 49-59; February, 1957.

⁷ Having discussed the role of the higher modes in the slots, it has to be pointed out that they are negligible because of (2). This can be checked by the position of the shorting plane in a metal plate medium with plate thickness $D_2 = \text{zero}$. For instance for $\lambda = 10D_1$ and $\epsilon_i = \epsilon_0$, it lies at $l/D_1 = 0.22$ from the edge of the medium. See Marcuvitz, "Waveguide Handbook," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 289-292; 1951.

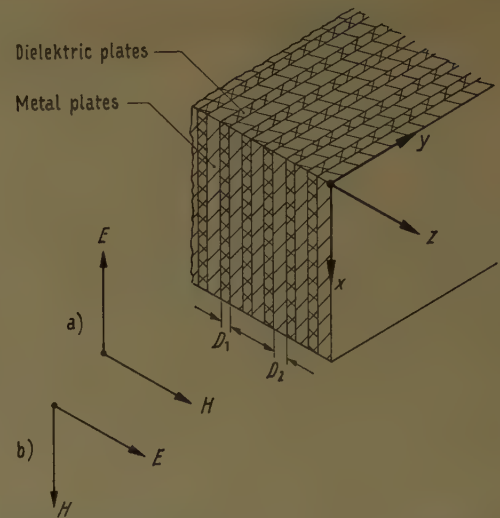


Fig. 2—Anisotropic medium in the Cartesian coordinate system with incident plane waves (schematically).

In this case, permeability of the dielectric and the metal equals that of space. For the permeability perpendicularly to the plates there holds, thus,

$$\mu_z = \mu_0. \quad (6)$$

If, however, a plane wave with the electrical field *perpendicularly* to the plates strikes the medium in the direction y [see Fig. 2(b)], a plane wave can propagate between the metal plates. Because of (3), the plates are here virtually field-free. Perpendicularly to the plates the medium, therefore, represents a dielectric with the dielectric constant

$$\epsilon_z = \epsilon_i \frac{D_1 + D_2}{D_1}. \quad (7)$$

The permeability in parallel to the plates is

$$\mu_p = \mu_0 \frac{D_1}{D_1 + D_2}. \quad (8)$$

The cylinder-symmetrical medium 1 thus presents in a radial and a circular direction the dielectric constant ϵ_p of (4) and (5) and the permeability μ_p of (8). In an axial direction the medium 1 has the dielectric constant ϵ_z of (7) and the permeability μ_0 of space.

The tensors of dielectric constant ϵ_1 and permeability μ_1 in medium 1 thus have the following form

$$\epsilon_1 = \begin{pmatrix} \epsilon_p & 0 & 0 \\ 0 & \epsilon_p & 0 \\ 0 & 0 & \epsilon_z \end{pmatrix} \quad (9)$$

$$\mu_1 = \begin{pmatrix} \mu_p & 0 & 0 \\ 0 & \mu_p & 0 \\ 0 & 0 & \mu_0 \end{pmatrix}. \quad (10)$$

The Propagation Constants of the Individual Modes

General Considerations: Pure H-modes and E-modes exist also, in this case, only with rotation-symmetrical fields, just as in the case of circular waveguides. In the case of not-rotation-symmetrical fields, an H-mode must always be combined with an E-mode in order to satisfy the boundary conditions. The rotation-symmetrical modes are, therefore, here termed H_0 and E_0 modes, and the not-rotation-symmetrical modes are called HE_m -modes (H-type prevailing) and EH_m -modes (E-type prevailing). Corresponding to the homogeneous wire guide, only the E_0 -mode can have low attenuation also with the cylindrical corrugated guide. Its phase velocity then is but slightly lower than the velocity of light. The H_0 , HE_m , and EH_m -modes have very high attenuation and are of no practical interest. They correspond to the spurious modes in Sommerfeld.¹

The plane corrugated guide has no H_0 and HE_m modes at all, only E_0 and EH_m modes.

If the frequency is sufficiently high, the cylindrical corrugated guide corresponds to the plane corrugated guide, and the propagation constants of the EH_m -modes differ but slightly from that of the E_0 -mode (see *Solutions for $|k_0 b| \gg \frac{1}{2}$*).

With a bare copper wire, i.e., $d=0$, for instance, there is, with a diameter of 10 cm, the difference of 10 per cent between the E_0 and EH_1 modes in the quantities k_0 if the wavelength is $\lambda_0=0.15$ mm.

The following discussion is restricted to the E_0 mode, because it alone holds practical interest.

Phase and Attenuation Constants of the Rotation-Symmetrical E_0 -Mode with $\epsilon_i=\epsilon_0$ and Small b/λ_0 : From (42) to (49) the following formulas result, in this case, for the phase and attenuation constants.

$$\beta = \beta_0 \sqrt{1 + \left(\frac{\lambda_0}{2\pi r_0}\right)^2} \quad (11)$$

$$\alpha = \frac{R}{2} \frac{1 + \frac{2a}{D_1 + D_2} \ln \frac{b}{a}}{60 \ln \frac{0.68r_0}{b}} \quad (12)$$

$$\text{for } \frac{2\pi b}{\lambda_0} \ll 1.$$

Here

$$R = \frac{1}{2\pi a} \sqrt{\frac{\beta_0 Z_0}{2\kappa}} \quad (13)$$

denotes the resistance per unity length of the guide without corrugations. The quantity r_0 is calculated from the formula

$$\ln \frac{1.123r_0}{b} = \left(\frac{2\pi r_0}{\lambda_0}\right)^2 \frac{D_1}{D_1 + D_2} \ln \frac{b}{a}. \quad (14)$$

From (12) there results the special case

$$\alpha = \frac{R}{60 \ln \frac{0.68r_0}{b}} \quad (15)$$

$$\text{for } \frac{d}{a} \ll 1; \quad D_1 + D_2 = 2d.$$

The factor in (12) multiplied by $R/2$ can be considered as the characteristic impedance Z of the guide.

Fig. 3 shows, for $D_1=D_2$ and $\epsilon_i=\epsilon_0$, the quantity b/r_0 as a function of $b/\lambda_0 \sqrt{\log_e (b/a)}$. The significance of the radius r_0 , which was introduced initially in Kaden,⁸ resides in the fact that more than 90 per cent of the wave energy travels in the space between $r=0$ and $r=r_0$.

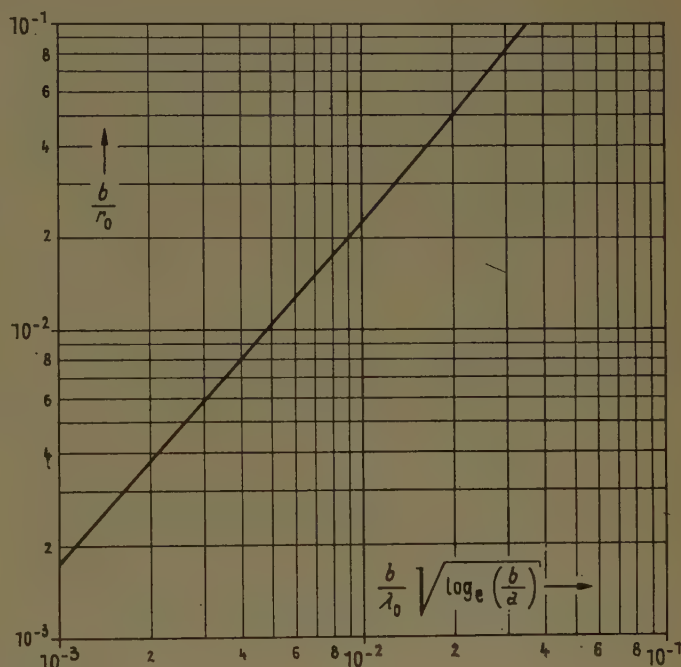


Fig. 3— b/r_0 for the cylindrical corrugated guide as a function of

$$\left(\frac{b}{\lambda_0}\right) \sqrt{\log_e \left(\frac{b}{a}\right)} \text{ with } D_1 = D_2 \text{ and } \epsilon_i = \epsilon_0.$$

Fig. 4 shows, on a cylindrical corrugated copper guide, the attenuation α and the radius r_0 as a function of wavelength (100 to 400 cm) for different b/a and D_1+D_2 , but invariably $D_1=D_2$ and $a=0.75$ cm. It is seen that with increasing wavelength the attenuation decreases and the radius r_0 increases. It is also evident, that with increasing b/a , i.e., increasing depth of the corrugations, the attenuation rises and r_0 falls correspondingly. It is interesting that with the constant b/a

⁸ H. Kaden, "Fortschritte in der Theorie der Drahtwellen," *Arch. elektr. Übertrag.*, vol. 5, pp. 399-414; September, 1951.

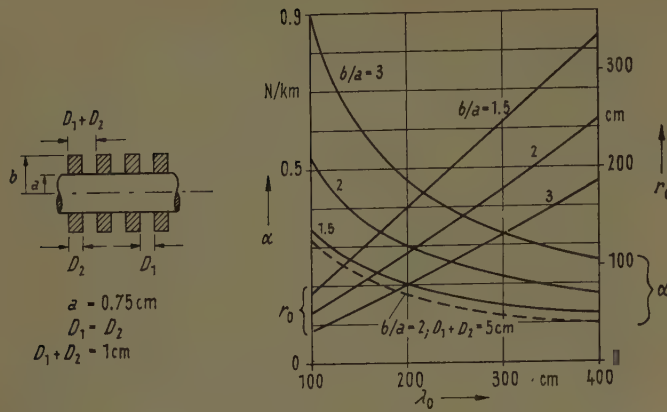


Fig. 4—The attenuation α and the radius r_0 as a function of the wavelength λ_0 with a cylindrical corrugated copper guide.

the attenuation falls with increasing D_1+D_2 as the broken curve shows. The radius r_0 , however, remains the same, for, according to (14), it depends only on the ratio $(D_1+D_2)/D_1$. This decrease of the attenuation is due to the fact that the path of the current along the guide decreases. In reality, however, the radius r_0 will slightly increase as well when D_1+D_2 goes up, and the attenuation will not decrease so heavily, for the ratio of wavelength to corrugation constant falls with increasing D_1+D_2 so that the inequality (1) is less well met.

For comparison, Fig. 5 shows the attenuation α and the radius r_0 with a Goubau guide of copper. The dielectric coating has a relative dielectric constant $\epsilon/\epsilon_0 = 2.5(1-j4 \cdot 10^{-4})$. The curves are taken from Piefke.⁹ The corresponding trend of the curve is obvious. For $b/a=2$ the curves of the Goubau guide agree approximately with the broken curve of Fig. 4 and the corresponding curve for r_0 .

Fig. 6 shows, for the cylindrical corrugated guide also, the phase velocity as a function of λ_0 . The dimensions are the same as in Fig. 4. It is seen that it differs from the velocity of light by but a few per cent, which difference, of course, increases with rising ratio b/a .

Phase and Attenuation Constants of the E-Mode with the Plane Corrugated Guide: With a plane corrugated guide the calculations are simpler so that the fundamental effect of corrugations can be well studied.

At a fixed frequency the locus of k_0/β_0 as a function of the corrugation depth d approximates a circle in the complex plane, as Fig. 7 shows for an example. Because of the high conductivity the circle passes here almost through the origin of the coordinate system. Most values d/λ_0 are near that point. The values about $d/\lambda_0=0.25$ thus cover most of the circle. For $0.25-d/\lambda_0 \ll 1$ there is $k_0/\beta_0 \gg 1$, and thus the propagation constant $\gamma \approx \pm k_0$. For $0.25 < d/\lambda_0 < 0.5$ no solution exists, because k_0 has here a negative imaginary component (the limits are here not exactly at $d/\lambda_0=0.25$ and $d/\lambda_0=0.5$, for the conductivity is finite).

⁹ G. Piefke, "Zur Theorie der Harms-Goubau-Drahtwellenleitung beim Meterwellen," *Arch. elekt. Übertrag.*, vol. 9, pp. 81-93; February, 1955.

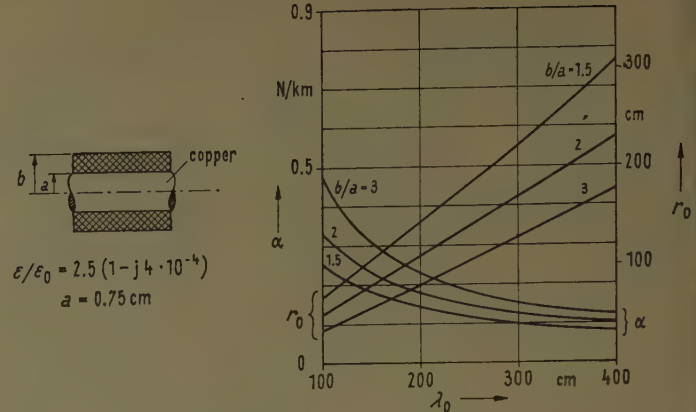


Fig. 5—The attenuation α and the radius r_0 of a Goubau guide.

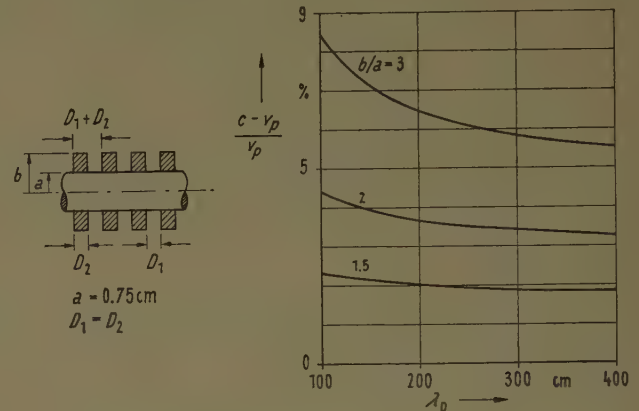


Fig. 6—The phase velocity v_p as a function of the wavelength λ_0 with a cylindrical corrugated guide.

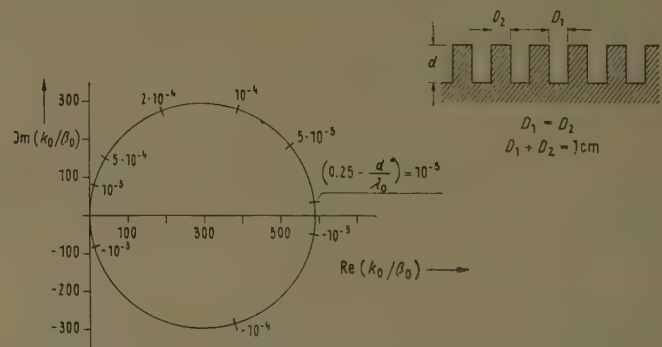


Fig. 7— k_0/β_0 with a plane corrugated copper guide as a function of d/λ_0 for $\lambda_0=200$ cm.

With the exception of the phase and attenuation constants for values d/λ_0 near $d/\lambda_0=k \cdot 0.25$ ($k=0, 1, 2, \dots$), α and β are available from the formulas

$$\beta = \beta_0 \sqrt{1 + \left(\frac{D_1}{D_1 + D_2} \tan \beta_0 d \right)^2} \quad (16)$$

$$\alpha = \frac{\beta_0 \sqrt{\frac{\beta_0}{2Z_0 k} \left(1 + \frac{d}{D_1 + D_2} \right)}}{\cos^2 \beta_0 d \sqrt{1 + \left(\frac{D_1 + D_2}{D_1} \cot \beta_0 d \right)^2}} \quad (17)$$

according to (54) to (57) which follow. Fig. 8 shows an evaluation of (17) for $D_1 = D_2$ and $D_1 + D_2 = 1$ cm. With the full curve there is here $\lambda_0 = 200$ cm, and d is variable. With the broken curve there is $d = 20$ cm, and λ_0 is variable. While for $d = 0$ the attenuation is very low (Sommerfeld wire), it becomes infinite near $d/\lambda_0 = 0.25$. For $0.25 \leq d/\lambda_0 \leq 0.5$ no solution exists of the wave equation. For $0.5 < d/\lambda_0 < 0.75$ solutions exist again, for $0.75 < d/\lambda_0 < 1$ there exist none, etc. The solutions in the range $0.5 < d/\lambda_0 < 0.75$ correspond to those in the range $0 < d/\lambda_0 < 0.25$, where, of course, the attenuation values are lower because of the lower corrugation depth. As mentioned above, the limits are not accurately at $d/\lambda_0 = k \cdot 0.25$ ($k = 0, 1, 2 \dots$) since the conductivity is finite (see Fig. 7).

The corrugated guide thus acts as a band-pass, with pass bands about at $k \cdot 0.5 < d/\lambda_0 < (0.25 + k \cdot 0.5)$ and stop bands about at $(0.25 + k \cdot 0.5) < d/\lambda_0 < 0.5(1 + k)$.

For the more accurate formulas for the propagation constants of the various modes refer to the following section where the problem is discussed mathematically.

MATHEMATICAL ANALYSIS OF THE PROBLEM

The Field Components in the Cylindrical Corrugated Guide

The Field Components for $0 \leq r \leq a$ (Medium 2): As Piefke¹⁰ shows, the field components in medium 2 are

$$E_z = C_1 \left\{ \begin{matrix} \sin m\phi \\ \cos m\phi \end{matrix} \right\} J_m(k_2 r)$$

$$H_z = C_2 \left\{ \begin{matrix} \cos m\phi \\ \sin m\phi \end{matrix} \right\} J_m(k_2 r)$$

$$H_\phi = - \left\{ \begin{matrix} \sin m\phi \\ \cos m\phi \end{matrix} \right\} \left[C_1 \frac{j\omega\epsilon_2}{k_2} J_m'(k_2 r) \pm C_2 \frac{m}{r} \frac{\gamma}{k_2^2} J_m(k_2 r) \right]$$

$$E_\phi = \left\{ \begin{matrix} \cos m\phi \\ \sin m\phi \end{matrix} \right\} \left[\pm C_1 \frac{m}{r} \frac{\gamma}{k_2^2} J_m(k_2 r) + C_2 \frac{j\omega\mu_2}{k_2} J_m'(k_2 r) \right] \quad m = 0, 1, 2, \dots \quad (18)$$

$$k_2 = \sqrt{\omega^2 \mu_2 \epsilon_2 + \gamma^2} = \sqrt{\beta_0^2 \frac{\mu_2 \epsilon_2}{\mu_0 \epsilon_0} + \gamma^2} \approx (1 - j) \sqrt{\frac{\omega \mu_0 \kappa}{2}}, \quad (19)$$

since with metal (e.g., copper) $\mu_2 = \mu_0$ and $\epsilon_2 = \kappa/(j\omega)$. All field components are proportional to $e^{\gamma z} e^{j\omega t}$. These factors have been dropped in (18). The radial field components are not stated because they are not needed to satisfy the boundary conditions. In (18), C_1 and C_2

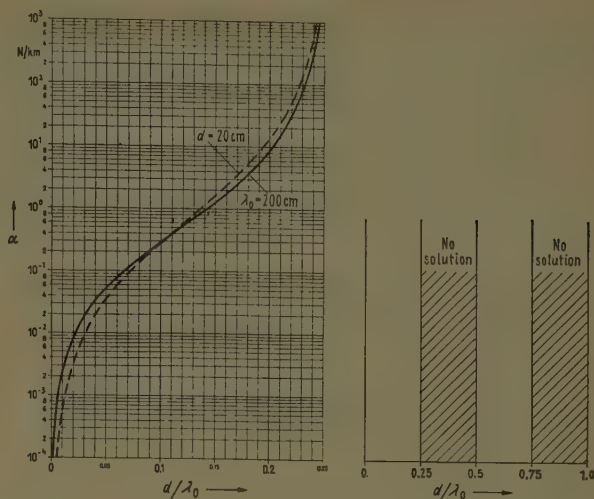


Fig. 8—The attenuation α with a plane corrugated copper guide as a function of the corrugation depth d and the wavelength λ_0 for $D_1 = D_2$, $D_1 + D_2 = 1$ cm.

are constants, C_1 corresponding to the amplitude of the E-mode and C_2 to that of the H-mode. The upper sign in (18) refers to the upper trigonometric function, and the lower sign to the lower trigonometric function; e.g., the positive sign preceding C_2 relates in the case of field strength H_ϕ to $\sin m\phi$ and the negative sign to $\cos m\phi$.

The Field Components for $a \ll r \ll b$ (Medium 1): Because of (4) to (10), and corresponding to Piefke,¹⁰ the field components here are

$$E_z = \left\{ \begin{matrix} \sin m\phi \\ \cos m\phi \end{matrix} \right\} [T_1 H_m^{(2)}(k_E r) + R_1 H_m^{(1)}(k_E r)]$$

$$H_z = \left\{ \begin{matrix} \cos m\phi \\ \sin m\phi \end{matrix} \right\} [T_2 H_m^{(2)}(k_H r) + R_2 H_m^{(1)}(k_H r)]$$

$$H_\phi = - \left\{ \begin{matrix} \sin m\phi \\ \cos m\phi \end{matrix} \right\} \left[\frac{j\omega\epsilon_s}{k_E} (T_1 H_m^{(2)'}(k_E r) + R_1 H_m^{(1)'}(k_E r)) \right]$$

$$\pm \frac{m}{r} \frac{\gamma \mu_s}{k_H^2 \mu_p} (T_2 H_m^{(2)}(k_H r) + R_2 H_m^{(1)}(k_H r))$$

$$E_\phi = \left\{ \begin{matrix} \cos m\phi \\ \sin m\phi \end{matrix} \right\} \left[\pm \frac{m}{r} \frac{\gamma \epsilon_s}{k_E^2 \epsilon_p} (T_1 H_m^{(2)}(k_E r) + R_1 H_m^{(1)}(k_E r)) \right]$$

$$+ \frac{j\omega\mu_s}{k_H} (T_2 H_m^{(2)'}(k_H r) + R_2 H_m^{(1)'}(k_H r)) \quad (20)$$

$$m = 0, 1, 2, \dots$$

$$k_E = \sqrt{\frac{\epsilon_s}{\epsilon_p} (\omega^2 \mu_p \epsilon_p + \gamma^2) - j\alpha_E} \approx \beta_0 \sqrt{\frac{\epsilon_s}{\epsilon_0}} - j \frac{1}{Z_0 D_1} \sqrt{\frac{\omega \mu_0}{2\kappa}} \quad (21)$$

$$k_H = \sqrt{\frac{\mu_s}{\mu_p} (\omega^2 \mu_p \epsilon_p + \gamma^2)} \approx k_2 \sqrt{\frac{D_2}{D_1 + D_2}} \quad (22)$$

¹⁰ G. Piefke, "Die Übertragungseigenschaften einer Leitung aus axial angeordneten, voneinander isolierten Metallringen," *Arch. elektr. Übertrag.*, vol. 11, pp. 423-428 and 449-454; October, 1957.

where T_1 , T_2 , R_1 , and R_2 are constants that must be determined from the boundary conditions. The quantity α_E results from the losses of the metal in the space $a < r < b$. The term $(1/Z_0 D_1) \sqrt{\omega \mu_0 / 2\kappa}$ is the attenuation constant of a plane wave in a parallel-strip transmission line with the spacing D_1 and the conductivity κ . The upper sign in (20) once more relates to the upper, the lower sign to the lower, trigonometric function.

The Field Components for $b \leq r \leq \infty$ (Medium 0): In the medium 0 the field components read

$$\begin{aligned} E_z &= K_1 \begin{Bmatrix} \sin m\phi \\ \cos m\phi \end{Bmatrix} H_m^{(1)}(k_0 r) \\ H_z &= K_2 \begin{Bmatrix} \cos m\phi \\ \sin m\phi \end{Bmatrix} H_m^{(1)}(k_0 r) \\ H_\phi &= - \begin{Bmatrix} \sin m\phi \\ \cos m\phi \end{Bmatrix} \left[K_1 \frac{j\omega\epsilon_0}{k_0} H_m^{(1)'}(k_0 r) \right. \\ &\quad \left. \pm K_2 \frac{m}{r} \frac{\gamma}{k_0^2} H_m^{(1)}(k_0 r) \right] \\ E_\phi &= \begin{Bmatrix} \cos m\phi \\ \sin m\phi \end{Bmatrix} \left[\pm K_1 \frac{m}{r} \frac{\gamma}{k_0^2} H_m^{(1)}(k_0 r) \right. \\ &\quad \left. + K_2 \frac{j\omega\mu_0}{k_0} H_m^{(1)'}(k_0 r) \right] \quad (23) \\ m &= 0, 1, 2, \dots \end{aligned}$$

$$k_0 = \sqrt{\omega^2 \mu_0 \epsilon_0 + \gamma^2} = \sqrt{\beta_0^2 + \gamma^2}. \quad (24)$$

The imaginary part of the quantity k_0 must always be positive.

The Equation for the Eigenvalue $k_0 b$

Because of the high conductivity κ of the guide, the asymptotic approximations are inserted for the Bessel functions with the argument $k_2 a$ and the Hankel functions with the argument $k_H r$. It is also assumed that the quantity $d = b - a$ is much greater than the equivalent thickness of the conducting layer of the metal. The H-mode traveling from $r = a$ in the direction r with the radial propagation constant k_H will then have practically vanished at the point $r = b$, and that traveling from $r = b$ in the direction towards negative r will have equally disappeared at the point $r = a$. There holds thus,

$$\begin{aligned} H_m^{(2)}(k_H b) &= H_m^{(2)'}(k_H b) = H_m^{(1)}(k_H a) \\ &= H_m^{(1)'}(k_H a) = 0. \end{aligned} \quad (25)$$

By satisfying the continuity conditions at the points $r = a$ and $r = b$, and neglecting higher-order terms in $1/k_2$ and $1/k_H$, one then obtains

$$\begin{aligned} \left(\frac{m\gamma}{k_0^2 b} \right)^2 &= - \left[\frac{\beta_0}{k_0} \frac{H_m^{(1)'}(k_0 b)}{H_m^{(1)}(k_0 b)} - j \frac{Z_H}{Z_0} \right] \\ &\quad \cdot \left[\frac{\beta_0}{k_0} \frac{H_m^{(1)'}(k_0 b)}{H_m^{(1)}(k_0 b)} - j \frac{Z_0}{Z_i} \right] \\ m &= 0, 1, 2, \dots \quad (26) \end{aligned}$$

where

$$\frac{Z_i}{Z_0} = j \frac{k_E \epsilon_0}{\beta_0 \epsilon_s} \frac{\frac{\partial F}{\partial(k_E a)} - j \frac{k_E \epsilon_0}{\beta_0 \epsilon_s} \frac{Z_0}{Z_I} F}{\frac{\partial^2 F}{\partial(k_E a) \partial(k_E b)} - j \frac{k_E \epsilon_0}{\beta_0 \epsilon_s} \frac{Z_0}{Z_I} \frac{\partial F}{\partial(k_E b)}} \quad (27)$$

$$F = J_m(k_E a) N_m(k_E b) - N_m(k_E a) J_m(k_E b) \quad m = 0, 1, 2, \dots \quad (28)$$

$$\frac{Z_I}{Z_0} = \sqrt{\frac{\epsilon_0}{\epsilon_s}} = (1 + j) \sqrt{\frac{\beta_0}{2Z_0 \kappa}} \quad (29)$$

$$\frac{Z_H}{Z_0} = \frac{Z_I}{Z_0} \sqrt{\frac{D_1 + D_2}{D_2}} \quad (30)$$

With $D_1 = 0$ and $d = 0$, i.e., $Z_H = Z_i = Z_I$, (26) to (30) yield the equation derived in Hondros² for a single wire if the high-argument approximations are inserted, in that paper, for the Bessel functions with the argument corresponding to the quantity $k_2 a$, and if the term corresponding to the quantity $1/k_2^2$ is neglected with respect to that corresponding to the quantity $1/k_0^2$.

With $m = 0$, and neglecting the losses, i.e., $Z_I = 0$, there results from (26) to (28) the equation mentioned in Rotman.⁵

For numerical evaluation it is advisable to replace the quantity Z_i/Z_0 by approximations. With insertion of the large-argument approximations of the Bessel and Neumann functions there results, thus, from (27) and (28) with $d = b - a$

$$\begin{aligned} \frac{Z_i}{Z_0} &= \frac{Z_I}{Z_0} \frac{1 + j \frac{k_E \epsilon_0}{\beta_0 \epsilon_s} \frac{Z_0}{Z_I} \tan k_E d}{1 + j \frac{\beta_0 \epsilon_s}{k_E \epsilon_0} \frac{Z_I}{Z_0} \tan k_E d} \\ &\approx \frac{Z_I}{Z_0} \frac{1 + j \frac{Z_E}{Z_I} \tan k_E d}{1 + j \frac{Z_I}{Z_E} \tan k_E d} \\ &\text{for } k_E a \gg \begin{cases} 1 \\ m \end{cases} \end{aligned} \quad (31)$$

where

$$\frac{Z_E}{Z_0} = \sqrt{\frac{\epsilon_0}{\epsilon_s}} \frac{D_1}{D_1 + D_2} \quad (32)$$

In the same way, (27) and (28) yield, with insertion of the small-argument approximations of the Bessel and Neumann functions,

$$\frac{Z_i}{Z_0} = \frac{\frac{Z_I}{Z_0} \left(\frac{1}{k_E a} - \frac{k_E a}{2} \ln \frac{1.123}{k_E b} \right) + j \frac{k_E \epsilon_0}{\beta_0 \epsilon_z} \ln \frac{b}{a}}{\frac{1}{k_E b} - \frac{k_E b}{2} \ln \frac{1.123}{k_E a} + j \frac{1}{2} \frac{Z_I}{Z_0} \frac{\beta_0 \epsilon_z}{k_E \epsilon_0} \left(\frac{b}{a} - \frac{a}{b} \right)}$$

$$\approx \frac{\frac{Z_I}{Z_0} \left(\frac{1}{k_E a} - \frac{k_E a}{2} \ln \frac{1.123}{k_E b} \right) + j \frac{k_E \epsilon_0}{\beta_0 \epsilon_z} \ln \frac{b}{a}}{\frac{1}{k_E b} - \frac{k_E b}{2} \ln \frac{1.123}{k_E a}}$$

$$m = 0, k_E b \ll 1. \quad (33)$$

Particular Solutions of the Equation for the Eigenvalue $k_0 b$

Solutions for $|k_0 b| \ll 1$: Insertion of the equations of the Hankel functions for small arguments in (26) yields

$$\frac{k_0}{\beta_0} \frac{H_0^{(1)}(k_0 b)}{H_0^{(1)'}(k_0 b)} \approx \frac{(k_0 b)^2}{\beta_0 b} \ln \frac{1.123 j}{k_0 b} = j \frac{Z_0}{Z_H} \quad (34)$$

for $m=0$, H_0 modes;

$$\frac{(k_0 b)^2}{\beta_0 b} \ln \frac{1.123 j}{k_0 b} = j \frac{Z_i}{Z_0} \quad (35)$$

for $m=0$, E_0 modes;

$$(k_0 b)^2 = m^2 \frac{Z_i}{Z_H} + j m \beta_0 b \left(\frac{Z_0}{Z_H} + \frac{Z_i}{Z_0} \right) \quad (36)$$

for $m=1, 2, 3, \dots$, HE_m - and EH_m -modes

Because of the very large Z_0/Z_H value, (34) and (36) yield only solutions at very low frequencies. Such solutions are at variance, however, with the approximations in (19), (21), and (22). Since (26) is based on these assumptions, the solutions gained from (34) and (36) are impractical. This means that with $|k_0 b| \ll 1$ no H_0 -, HE_m - and EH_m -modes result with $v_p \approx c$ and low attenuation. With the E_0 -modes however, results with $|k_0 b| \ll 1$ and $\beta_0 \gg |k_0|$ exist perfectly well, as (35) shows. This means that the E_0 -modes can travel with approximately the velocity of light and with low attenuation.^{1,2}

Solutions for $|k_0 b| \gg 1$: After insertion of the approximations of the large-argument Hankel functions into (26), there results

$$\frac{k_0}{\beta_0} = \frac{Z_0}{Z_H} \quad (37)$$

for $m=0$, H_0 -modes, $m \neq 0$, HE_m -modes, $b \rightarrow \infty$;

$$\frac{k_0}{\beta_0} = \frac{Z_i}{Z_0} \quad (38)$$

for $m=0$, E_0 -modes, $b \rightarrow \infty$;

$$\frac{k_0}{\beta_0} = \frac{Z_i}{Z_0} (1 + \eta) \quad (39)$$

for $m=0, 1, 2, \dots$, EH_m -modes, $\eta \ll 1$, $b \rightarrow \infty$;

$$\eta = \left(\frac{m}{\beta_0 b} \right)^2 \frac{\left(\frac{Z_0}{Z_i} \right)^2 - 1}{1 + 2 \left(\frac{m}{\beta_0 b} \right)^2 - \frac{Z_H Z_i}{Z_0^2}} \approx \left(\frac{m}{\beta_0 b} \frac{Z_0}{Z_i} \right)^2$$

for $\left| \frac{Z_0}{Z_i} \right| \gg 1. \quad (40)$

For the wire without corrugations, there results

$$\eta = \left(\frac{m}{\beta_0 b} \frac{Z_0}{Z_i} \right)^2. \quad (41)$$

A solution according to (37) is not possible since the quantity Z_0/Z_H always has a negative imaginary part because of (29) and (30). No H_0 - and HE_m -modes exist, thus, with the plane corrugated guide.^{11,12}

Eqs. (38) to (41) show that E_0 -modes exist on the plane corrugated guide and that with very large $\beta_0 b$ the propagation constant of the EH_m -modes agrees approximately with the propagation constant of the E_0 -modes. With the solution as per (38) to (41), it should always be noted that the approximations of (19), (21), and (22) must be satisfied and that k_0 must always have a positive imaginary part.

Insertion of (31) in (38) yields, with $Z_I=0$, the equation for the plane problem mentioned in Rotman.⁵

Calculation of the Propagation Constant γ

The Propagation Constant with $m=0$, $\epsilon_i=\epsilon_0$, $|k_0 b| \ll 1$, and $\beta_0 b \ll 1$: Eq. (24) yields

$$\gamma = \pm j(\beta - j\alpha) = \pm j\beta_0 \sqrt{1 - \left(\frac{k_0}{\beta_0} \right)^2}. \quad (42)$$

With

$$\left| \frac{k_0}{\beta_0} \right|^2 \ll 1 \quad (43)$$

and

$$k_0 = j \frac{1}{r_0} (1 - j\xi), \quad \xi \ll 1, \quad (44), (45)$$

¹¹ W. O. Schumann, "Elektrische Wellen," Carl Hanser Verlag, München, Germany, pp. 231-232; 1948.

¹² H. Kaden, "Dielektrische und metallische Wellenleiter," Arch. elektr. Übertrag., vol. 6, pp. 319-332; August, 1952.

(42) yields

$$\beta = \beta_0 \sqrt{1 + \left(\frac{\lambda_0}{2\pi r_0}\right)^2} \quad (46)$$

$$\alpha = \frac{\xi}{\beta_0 r_0^2} \quad (47)$$

Eqs. (21) and (29) are now inserted in (33). Subsequently, (33) and (44) are inserted in (35). Because of the high conductivity κ and (45), the following equations for r_0 and ξ result, then, with consideration of (7):

$$\ln \frac{1.123 r_0}{b} = \left(\frac{2\pi r_0}{\lambda_0}\right)^2 \frac{D_1}{D_1 + D_2} \ln \frac{b}{a} \quad (48)$$

$$\xi = \frac{1}{2a} \sqrt{\frac{\beta_0}{2Z_0\kappa}} \frac{(\beta_0 r_0)^2}{\beta_0} \frac{1 + \frac{2a}{D_1 + D_2} \ln \frac{b}{a}}{\ln \frac{0.68 r_0}{b}} \quad (49)$$

The attenuation constant α can then, at once, be calculated from (47).

The Propagation Constant γ with $m=0$, $\epsilon_i=\epsilon_0$, $|k_0 b| \gg 1$, $\beta_0 b \gg 1$: Insertion of (31) into (38) yields

$$\frac{k_0}{\beta_0} = \frac{Z_I}{Z_0} \frac{1 + j \frac{Z_E}{Z_I} \tan k_E d}{1 + j \frac{Z_I}{Z_E} \tan k_E d} \quad (50)$$

Insertion of (21), (29), and (32) into (50) now gives

$$\frac{k_0}{\beta_0} = \frac{1 + \exp^{-j2 \arctan \Phi}}{2 \frac{D_1 + D_2}{D_1} \sqrt{\frac{\beta_0}{2Z_0\kappa}} \left(1 + \frac{d}{D_1}\right)} \quad (51)$$

for $\frac{\pi}{2} - \beta_0 d \ll 1$

where

$$\Phi = \frac{2\pi \left(0.25 - \frac{d}{\lambda_0}\right)}{1 - \frac{\sqrt{\frac{\beta_0}{2Z_0\kappa}}}{1 + \frac{d}{D_1}}} \quad (52)$$

There results, further,

$$\frac{k_0}{\beta_0} = \frac{\sqrt{\frac{\beta_0}{2Z_0\kappa}}}{\cos^2 \beta_0 d} \left(1 + \frac{d}{D_1 + D_2}\right) + j \left(\frac{D_1}{D_1 + D_2} \tan \beta_0 d + \sqrt{\frac{\beta_0}{2Z_0\kappa}}\right) \quad (53)$$

for $\left|\frac{D_1 + D_2}{D_1} \sqrt{\frac{\beta_0}{2Z_0\kappa}} \tan k_E d\right| \ll 1$

Under the assumptions

$$\left|\frac{D_1 + D_2}{D_1} \sqrt{\frac{\beta_0}{2Z_0\kappa}} \tan k_E d\right| \ll 1 \quad (54)$$

$$\left|\frac{D_1}{D_1 + D_2} \tan \beta_0 d\right| \gg \sqrt{\frac{\beta_0}{2Z_0\kappa}} \quad (55)$$

there results, upon insertion of (53) into (42),

$$\beta = \beta_0 \sqrt{1 + \left(\frac{D_1}{D_1 + D_2} \tan \beta_0 d\right)^2} \quad (56)$$

$$\alpha = \frac{\beta_0 \sqrt{\frac{\beta_0}{2Z_0\kappa}} \left(1 + \frac{d}{D_1 + D_2}\right)}{\cos^2 \beta_0 d \sqrt{1 + \left(\frac{D_1 + D_2}{D_1} \cot \beta_0 d\right)^2}} \quad (57)$$

Radiation and Guided Waves

A. E. KARBOWIAK†

Summary—An analysis is carried out of the field radiated by a dipole in an arbitrary medium separated from others by parallel plane interfaces.

The resulting solution in the form of a contour integral is then examined in relation to the intrinsic properties of the media involved and the position of the antenna. The solution is obtained by splitting the integral into pole residues and branch-cut-integral; the latter is evaluated by developing it into a suitable asymptotic series.

It is shown that depending on conditions modal-type propagation can take place between the parallel interfaces in addition to radiation field. Distinction is drawn between proper modes, quasi-modes, surface waves, leaky waves and radiation field. All these waves are needed in the complete description of the field, and their relative intensities are evaluated.

Extreme cases are considered in which the media involved pass into good conductors on one hand and perfect dielectrics on the other.

Surface waves appear as a particular case of the more general problem considered (or as a part of the solution) and they form just one piece of a complete jig-saw, though under certain specifically simple conditions these waves appear to have some rather unique properties. In general (with the exception of perfectly conducting tubular waveguides) it is shown that no mode can exist on its own but rather in conjunction with other modes (including surface waves) and the radiation field.

LIST OF PRINCIPAL SYMBOLS

- $\mu_i, \epsilon_i, \sigma_i$ = constants of the i th medium
 $k_i = \omega \sqrt{\mu_i \epsilon_i}$ = propagation coefficient of the i th medium
 Z_s = surface impedance
 γ = propagation coefficient of the l th mode
 K^i = dipole intensity
 h_l = cutoff coefficient of the l th mode
 ψ_n = n th pole wave
 d = separation between the waveguide plates
 t = height of the dipole above the ground plane
 u = decay coefficient
 $Z_i = \sqrt{\mu_i / \epsilon_i}$ = intrinsic impedance of the i th medium
 $Z = (Z_1 / Z_2) (k_2 / k_1)$
 a, g, b = coefficients
 Λ = a quantity defined by (40)–(45)
 R_i = residue at the i th pole

INTRODUCTION

FROM the earliest days in the history of electromagnetic theory it has been customary to draw distinction between "free waves" and "bound" or "guided waves." Yet, from among the great wealth of the written information, it is impossible to extract a clear distinction between the two types of waves. Free waves are thought to be electromagnetic waves which spread out, from a source, in a medium of infinite extent. Guided waves, on the other hand, are associated with boundaries which confine or channel the electromagnetic energy in accordance with the properties and

the manner in which the boundaries are distributed. Such a distinction, though adequate in many cases, is confusing in others.

The early work of Sommerfeld¹ is a good illustration of the problems involved. In its original formulation, the problem is to find an expression for the field at a long distance from a Hertzian dipole situated at a certain height above a plane earth. The formal solution to the problem (in the form of a contour integral) presents no fundamental difficulty and the integral expression obtained is, mathematically, certainly correct; the difficulty lies elsewhere. The closed form, a rigorous solution in the form of an integral, is devoid of physical meaning and is incapable of useful and direct physical interpretation.

A more difficult part of the problem is now the interpretation of the closed-form solution, using suitable approximations, and it is precisely from this point onwards that the opinions, methods of approach and interpretation of the results differ, as obtained by various investigators. Sommerfeld himself obtained approximate solutions by deforming the path of integration and expanding the resulting integral in a suitable asymptotic series. The solution was composed of a "space wave" and a "surface wave." Whereas all concerned seem to have understood the meaning and physical interpretation of "space wave" (radiation), "surface wave" became a topic of lengthy discussions for decades.

Many investigators have since examined and re-examined the problem in detail,² but apart from a sign correction in the early Sommerfeld paper and a distinction drawn between "surface wave" and "ground wave," the question of existence of surface waves has not been settled.

More recently solutions have been obtained to a number of allied problems³ all concerned with the field produced by a dipole (point or line) located in the vicinity of loss-free structures. It transpires that with such

¹ A. Sommerfeld, "Über die Ausbreitung der Wellen in der drahtlosen Telegraphie," *Ann. d. Phys.*, vol. 28, p. 665; 1909.

² J. Zenneck, "Über die Fortpflanzung ebener elektromagnetischer Wellen langs einer ebenen Leiter fläche und ihre Beziehung zur drahtlosen Telegraphie," *Ann. d. Phys.*, vol. 23, p. 846; 1907.

W. H. Wise, "The physical reality of Zenneck's surface wave," *Bell Sys. Tech. J.*, vol. 16, pp. 35–44; 1957.

T. Kachan and G. Eckart, "Zur Frage der Oberflächen-Wellen in der Dipolestrahlung über einer ebenen Erde," *Arch. Elec. Übertragung*, vol. 5, pp. 347–348; 1951.

³ R. M. Whitmer, "Fields in Nonmetallic Waveguides," *Proc. IRE*, vol. 36, pp. 1105–1109; September, 1948.

C. T. Tai, "The effect of a grounded slab on the radiation from a line source," *J. Appl. Phys.*, vol. 22, p. 405; 1951.

A. L. Cullen, "The excitation of plane surface waves," *Proc. IEE Monograph No. 93R*; February, 1954.

† Standard Telecommunication Labs., London, England.

structures there is never any doubt whether the surface wave exists or not and quantitative results can be obtained.

Because of obvious mathematical difficulties "lossy" structures (apart from modifications of Sommerfeld's problem)⁴ have not been investigated in any detail. Yet, experimental evidence is such as to suggest that Zenneck wave does not exist,⁵ although surface waves over reactive surfaces are an established experimental fact.⁶ There is also ample experimental evidence in support of the existence of axial cylindrical surface wave (E_{0z}) first discussed by Sommerfeld⁷ and from some of the experiments obtained one should conclude that a Zenneck wave perhaps is a physical reality.⁸

At the same time, it is instructive to observe that doubts about the physical reality of guided waves in tubular metallic waveguides have never been raised, whether the waveguides be lossy or loss-less.

At this point, one cannot help wondering whether the whole question of existence and reality of guided waves is more a matter of definitions and the order of the magnitudes of the quantities involved, rather than distinct physical differences; this being the case, what is the right way of interpreting the mathematical expressions involved? The aim of this paper is to investigate this question and problems connected with it.

DISCUSSION OF THE PROBLEMS

The subdivision of a dynamic electromagnetic field into radiation field and guided waves is purely arbitrary and the only justification for this procedure lies in convenience: convenience of description, formulation of a clearer physical picture, etc. Although all electromagnetic waves are solutions to Maxwell's equations subject to the boundary conditions (and/or radiation condition at infinity) as well as launching conditions, the physical reality of any wave could be⁹ a matter of lengthy philosophical discussions.

⁴ H. G. Brooker and P. C. Clemmow, "A relation between the Sommerfeld theory of radio propagation over a flat earth and the theory of diffraction at a straight edge," *Proc. IEE*, vol. 97, pt. 3, p. 18; 1950.

J. R. Wait and W. C. R. Fraser, "Radiation from a vertical dipole over a stratified ground," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-1, pp. 9-11, July, 1953; Part II, *ibid.*, vol. AP-3, pp. 144-146, October, 1954.

⁶ C. R. Burrows, "The surface wave in radio propagation over plane earth," *PROC. IRE*, vol. 25, pp. 219-229; February, 1937.

⁷ H. M. Barlow and A. L. Cullen, "Surface waves," *Proc. IEE*, vol. 100, pt. 3, pp. 329-341; April, 1953.

H. M. Barlow and A. E. Karbowiak, "An experimental investigation of the properties of corrugated cylindrical surface waveguides," *Proc. IEE*, vol. 100, pt. 2, pp. 182-188; May, 1954.

⁷ G. Goubau, "Surface waves and their application to transmission lines," *J. Appl. Phys.*, vol. 21, p. 1119; 1950.

G. Goubau, "On the excitation of surface waves," *PROC. IRE*, vol. 40, pp. 865-868; July, 1952.

H. M. Barlow and A. E. Karbowiak, "An investigation of characteristics of cylindrical surface waves," *Proc. IEE*, vol. 100, pt. 3, pp. 321-328; November, 1953.

⁸ H. M. Barlow and A. E. Karbowiak, "An experimental investigation of axial cylindrical surface waves supported by capacitive surfaces," *Proc. IEE*, vol. 102-B, pp. 313-322; May, 1955.

⁹ Indeed the reality of some waves, notably the Zenneck wave, have been a subject of discussion for decades.

The reason for this confusion is as follows: On the one hand, it can be argued that it is a necessary but not sufficient condition for the existence of a wave-type that it shall be independently a solution to Maxwell's equations subject to boundary and launching conditions; but it is only when, in addition, it can be shown that such boundary conditions and launching devices can in fact be physically realized, that the wave may be said to exist. On the other hand, it can be argued that the wave-type investigated need not on its own satisfy Maxwell's equations, but so long as it forms a part of a field, and the latter satisfies Maxwell's equations, boundary and launching conditions, that the wave may be said to exist.

Surely, under these conditions when one is undecided on the meaning of the existence of a wave-type, before the problem of existence is tackled, the very question of existence cannot be answered in a unique and satisfactory manner, nor can a useful distinction between radiation and guided waves be made.

In an endeavor to determine what is useful (the only valid basis for definitions) we shall restrain from coining any definitions until we have reached some conclusions from the analysis of a number of representative problems.

In the problems to be investigated, we shall consider wave propagation from line dipole (Fig. 1) situated at a height t above a plane A separating regions 2 and 3. The region 2 is bounded by another plane B parallel to A and placed a distance d above it; the line dipole lies therefore parallel and between the planes A and B .

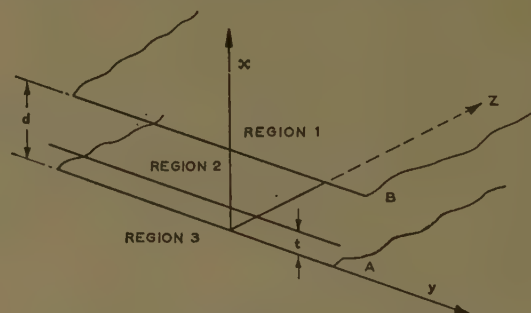


Fig. 1—The geometry of planar waveguide.

The co-ordinate axes are placed in such a way (Fig. 1) that the plane A contains axes z and y and the line dipole is parallel to y -axis at $z=0$.

The semi-infinite region 1 is assumed to be filled with a homogeneous medium 1 (constants $\mu_1, \epsilon_1, \sigma_1$); similarly region 2 is occupied by medium 2 (constants $\mu_2, \epsilon_2, \sigma_2$) and region 3 is occupied by medium 3 (constants $\mu_3, \epsilon_3, \sigma_3$).

In section 3, media 1 and 3 are assumed to be perfect conductors. In section 4, plane B is regarded as having a surface impedance¹⁰ Z_s and medium 3 a perfect conductor.

¹⁰ A. E. Karbowiak, "Theory of imperfect waveguides; the effect of wall impedance," *Proc. IEE*, vol. 102-B, pp. 698-708; September, 1955.

In section 5, the case of two different media filling regions 1 and 2 respectively are considered; medium 3 is assumed to be perfectly conducting.

A positive time factor $\exp(j\omega t)$ is implied in all relevant expressions, and the whole electromagnetic field (in the half space $z > 0$) is presumed to satisfy the radiation condition at infinity, since the only source of electromagnetic energy is the line dipole, situated close to the origin.

THE PERFECT CLOSED WAVEGUIDE

From the symmetry consideration, it is apparent that the only field components likely to be present, in the structure illustrated in Fig. 1 are E_x , E_z , and H_y . The elimination of the components of the electric field vector from the pair of curl equations of the Maxwell equation leads to the following inhomogeneous wave equation for the H_y component.¹¹

$$\frac{\partial^2 H_y}{\partial x^2} + \frac{\partial^2 H_y}{\partial z^2} + k_1^2 H_y = -jK\delta(z)\delta(x-t) \quad (1)$$

where

$k_1 = \omega\sqrt{\mu_1\epsilon_1}$ is the propagation coefficient

δ = Dirac's delta function

K = dipole intensity.

The introduction of the transform

$$H = \int_0^\infty H_y e^{-\gamma z} dz \quad (2)$$

into (1) permits an operational solution in the form for $0 < x < t$:

$$H = a_1 e^{ihx} + b_1 e^{-ihx} \quad (3)$$

$$H' = jh(a_1 e^{ihx} - b_1 e^{-ihx})$$

for $t < x < d$:

$$H = a_2 e^{ihx} + b_2 e^{-ihx} \quad (4)$$

$$H' = jh(a_2 e^{ihx} - b_2 e^{-ihx})$$

where the prime over H denotes differentiation with respect to x and where h is given by

$$h^2 = k_2^2 + \gamma^2. \quad (5)$$

Introducing the boundary conditions at $x=0$ and $x=d$ [$H'(0) = H'(d) = 0$] and observing the continuity of H at $x=t$ and that $H'(t+0) - H'(t-0) = -jK$, the solution for H can be obtained. In the range $t < x < d$, this is given by

$$H = -jK \frac{\cos ht \cdot \cos h(x-d)}{h \cdot \sin hd}. \quad (6)$$

The inversion of this expression yields the value of

¹¹ Note that all field components are normalized with respect to the characteristic impedance of the medium (Z_2).

H_y , thus

$$H_y = \frac{K}{2\pi} \int_{-j\infty}^{+j\infty} \frac{\cos ht \cdot \cos h(x-d)}{h \cdot \sin hd} e^{-\gamma z} d\gamma. \quad (7)$$

Evidently, the only singularities of the integrand are poles occurring for values of h given by

$$h_n = \frac{n\pi}{d} \quad (n = 0, \pm 1, \pm 2, \dots). \quad (8)$$

The solution to the integral (7) is, therefore, given by the sum of the residues at the poles, h_n , enclosed by the deformed path of integration as shown in Fig. 2.

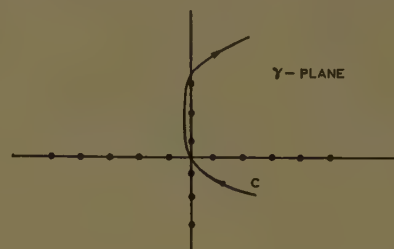


Fig. 2—Path of integration in the plane of γ , for a perfect waveguide.

To each and every value h_n , there correspond a propagation coefficient γ_n given by $\gamma_n = \sqrt{h_n^2 - k_2^2}$, the poles in the γ -plane. There is an infinity of these poles: some of the poles are situated on the imaginary axis, while all the remaining poles are distributed along the real axis. The integral for the entire field can be regarded, therefore, as an infinite sum of pole residues, whose amplitudes are given by

$$\psi_n = (H_y)_n = jK \frac{\cos h_n t \cdot \cos h_n(x-d)}{h_n} e^{-\gamma_n z} \quad (n \neq 0) \quad (9)$$

$$\psi_0 = \frac{jK}{2k_2 d} \cdot e^{-\gamma_0 z}$$

thus

$$H_y = \psi = \sum_{n=0}^{\infty} \psi_n. \quad (10)$$

We shall term the component waves ψ_n the pole-waves, since their existence is associated with the existence of the individual poles.

Clearly, whether a pole-wave exists or not depends on the height, t , of the dipole above the plane A : for values of t for which $\cos h_n t$ vanishes, the pole-wave vanishes, since the intensity of the pole waves is proportional to this factor. Furthermore, since the poles situated along the real axis give rise to evanescent waves, for $z \gg d$ these waves may be neglected because of their very small amplitude due to the factor $e^{-\gamma_n z}$. For large values of z the field is, therefore, almost entirely given by the sum of pole-waves having their origin in the finite number of poles distributed along the imaginary axis.

THE IMPERFECT WAVEGUIDE

If the plane B (Fig. 1) of the waveguide, analyzed in the last section, is replaced by a sheet having a surface impedance Z_s , then the field inside the waveguide will still be described by (3) and (4). The boundary conditions at $x=d$, however, must be changed to¹⁰

$$\left[\frac{H'}{H} \right]_{x=d} = jk_2 Z_s \quad (11)$$

where Z_s is the surface impedance (normalized with respect to Z_2) of the sheet B . With this change, the field inside the waveguide is given by [e.g., (7)]

$$H_y = \frac{K}{2\pi} \left[1 - \left(\frac{k_2}{h} Z_s \right)^2 \right]^{-1/2} \int_{-j\infty}^{+j\infty} \frac{\cos ht \cos h(x-d)}{h \sin(hd - \psi_s)} e^{-\gamma^2 d} d\gamma \quad (12)$$

where

$$\tan \psi_s = j \frac{k_2}{h} Z_s. \quad (13)$$

The integrand of (12) possesses an infinity of poles given as solutions of

$$d\sqrt{k_2^2 + \gamma^2} - \psi_s = n\pi \quad (14)$$

or

$$h_n = \sqrt{k_2^2 + \gamma_n^2} = \frac{n\pi}{d} + \frac{1}{d} \tan^{-1} \left(j \frac{k_2}{h} Z_s \right). \quad (15)$$

For small values of Z_s , the wave numbers h_n are given by a particularly simple expression

$$h_n \simeq \frac{n\pi}{d} + \frac{1}{d} \frac{k_2}{h} (jZ_s) = (h_n)_0 + \delta h. \quad (16)$$

That is, the wave numbers differ from those of a perfect waveguide ($n\pi/d$) by a small quantity

$$\delta h \left(= \frac{1}{d} \frac{k_2}{h} jZ_s \right).$$

The distribution of poles in relation to the path of integration is shown in Fig. 3. It will be observed that for small values of X_s , the solution is basically the same as for a perfect waveguide: the poles cluster close to the imaginary or real axis, although all poles are complex unless Z_s is purely imaginary.

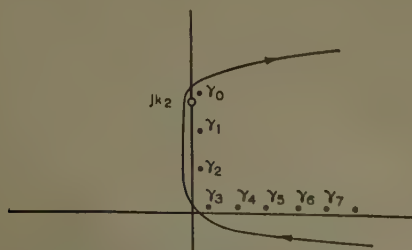


Fig. 3—Path of integration in the plane of γ , for an imperfect waveguide.

For convenience, we can, as with perfect waveguides, regard the field as composed of an infinity of pole waves of the form

$$\psi_n = A_n e^{-\gamma_n x}, \quad (17)$$

where γ_n denotes the poles which are given by

$$\gamma_n = \sqrt{h_n^2 - k_2^2} \quad (18)$$

$$\simeq \sqrt{\left(\frac{n\pi}{d} \right)^2 - k_2^2} + \delta\gamma_n, \quad (19)$$

$$= (\gamma_n)_0 + \delta\gamma,$$

with $\delta\gamma$ given by

$$\delta\gamma_n \simeq \frac{1}{d} \frac{k_2}{(\gamma_n)_0} (jZ_s). \quad (20)$$

The coefficients A_n [(17)] are, for small values of Z_s , given by (9) with h_n values given by (16).

It is important to observe that all poles lie in the first quadrant of γ -plane, and as with perfect waveguides, all poles (except γ_0) are situated below a line drawn parallel to the real axis and passing through the point jk_2 , the propagation coefficient of the medium 2. If the medium filling the waveguide is lossy, then k_1 will be complex (jk_1 will lie off the imaginary axis in the second octant of the Argand diagram). Accordingly, all $\delta\gamma$ values will be altered, but the pole-waves as defined above retain their meaning.

If $R_s(Z_s)$ is small, then all poles will be situated close to the imaginary axis (propagating waves) or close to the real axis (evanescent modes), though the "propagating waves" are slightly attenuated and the "evanescent waves" suffer a phase change with distance. The subdivision of waves into propagating and evanescent still retains a useful significance. As $R_s(Z_s)$ increases, however, the poles move further and further away from the real and imaginary axes, respectively, and the above subdivision of waves loses significance, though the concept of pole-waves as the sole constituents of the total field is valid. Moreover, it is still true that, because of their much higher attenuation, waves associated with poles situated further to the right of the origin become of lesser and lesser significance, as constituents of the total field, for greater and greater distances from the dipole.

WAVES IN PHYSICAL WAVEGUIDES

The Formal Solution to the Problem

If the regions 1 and 2 are filled with dissimilar homogeneous media 1 and 2, then the field in the region 2 is still given by (3) and (4), but these equations must be matched to the equations describing the field in region 1. This field is given by ($x > d$)

$$H = a_3 e^{-u(x-d)} \quad (21)$$

$$H' = -a_3 u e^{-u(x-d)},$$

The solution to the set of simultaneous equations (3), (4) and (21) subject to the boundary conditions of $x=0$, t and the continuity of the tangential field components at $x=d$ furnishes the following results for the coefficients

$$\begin{aligned} 2a_1 &= 2b_1 = W \left(\left(1 + Z \frac{j^v}{h} \right) e^{-j^v h(d-t)} \right. \\ &\quad \left. + \left(1 - Z \frac{j^v}{h} \right) e^{j^v h(d-t)} \right) \\ a_2 &= W \left(\left(1 + Z \frac{j^v}{h} \right) e^{-j^v h d} \cos ht \right) \\ b_2 &= W \left(\left(1 - Z \frac{j^v}{h} \right) e^{j^v h d} \cos ht \right) \\ a_3 &= W 2 \cos ht \end{aligned} \quad (22)$$

where

$$W = \frac{-K/h}{\left(1 + Z \frac{j^v}{h} \right) e^{-j^v h d} - \left(1 - Z \frac{j^v}{h} \right) e^{j^v h d}} \quad (23)$$

$$Z = \frac{Z_1}{Z_2} \frac{k_2}{k_1} \quad (24)$$

$Z_1 = \sqrt{\mu_1/\epsilon_1}$ = intrinsic impedance of medium 1, whose propagation coefficient is k_1

$Z_2 = \sqrt{\mu_2/\epsilon_2}$ = intrinsic impedance of medium 2, whose propagation coefficient is k_2 .

For the purpose of the study of the electromagnetic field, it is immaterial whether the field is examined in region 1 or 2. We propose to examine it in region 2, where ($0 < x < t$)

$$H = a_1 \cos hx. \quad (25)$$

Substituting for a , from (22) and inverting the expression (25), we obtain

$$H = \frac{K}{2\pi j} \int_{-j\infty}^{+j\infty} \frac{\cos hx}{h} \frac{\left(1 - Z \frac{j^v}{h} \right) e^{j^v h(d-t)} + \left(1 + Z \frac{j^v}{h} \right) e^{-j^v h(d-t)}}{\left(1 - Z \frac{j^v}{h} \right) e^{j^v h d} - \left(1 + Z \frac{j^v}{h} \right) e^{-j^v h d}} e^{-j^v x} d\gamma. \quad (26)$$

In this expression for the total field, the quantity Z is given by (24) and the wave numbers are connected by

$$\begin{aligned} k_1 &= \sqrt{-\gamma^2 - u^2} \\ k_2 &= \sqrt{-\gamma^2 + h^2}. \end{aligned} \quad (27)$$

Evidently, the integrand of (26) possesses branch points at $\gamma = \pm jk_1$ and an infinity of poles which are roots of the denominator of (26). When deforming the path of integration (from that running along the imaginary axis of γ) for the purpose of the evaluation of the

integral (26), it is essential to consider the position of the poles in relation to the branch cut and to choose the correct leaf of the Riemannian plane for the integration.

This branch of the complex γ -plane is named the "right Riemannian leaf," and the remaining parts the "wrong Riemannian leaf."

The right branch of γ is one over which u and γ are complex numbers in the first quadrant of the Argand diagram (to satisfy the radiation condition). The path of integration (C_0) indicated in Fig. 4 is a permissible one, provided that the branch point at $\gamma = jk_1$ and all the poles in the right leaf of the Riemannian plane to the right of the imaginary axes are enclosed between it and the infinity. Notwithstanding that the denominator of (26) possesses an infinite number of roots, the path C_0 may (as it usually does) enclose only a finite number of poles, the remaining poles being on the wrong Riemannian branch. The distribution of the poles and their position in relation to the branch-cut, which runs to the right of the origin and parallel to the real axes (Fig. 4), is a function of the constants of the media involved, and accordingly we shall differentiate between two somewhat distinct cases: 1) medium 2 denser than medium 1 [$|k_2| > |k_1|$]; 2) medium 1 denser than medium 2 [$|k_1| < |k_2|$].

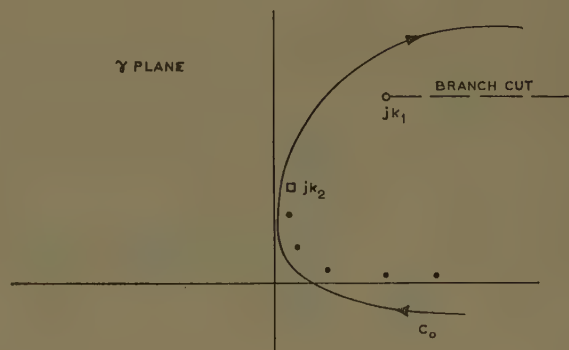


Fig. 4—A permissible path of integration (C_0) in the plane of γ .

To simplify matters, nonmagnetic media will be considered, so that $Z_2/Z_1 = k_1/k_2$.

Inside Medium 2 Denser Than the Outside Medium 1

Loss-Less Media: In general, the poles are given by the roots of the denominator of the expression (26) and the equations giving the roots can be put into the simple form

$$\begin{aligned} 1) \quad Z \frac{u}{h} &= \tan hd, \\ 2) \quad u &= \sqrt{-\gamma^2 - k_1^2}, \\ 3) \quad h &= \sqrt{\gamma^2 + k_2^2} \end{aligned} \quad (28)$$

which can be solved for h , thus

$$\left(\frac{k_2}{k_1} \right)^4 [(k_2^2 - k_1^2)d^2 - (hd)^2] = (hd)^2 \tan^2 hd. \quad (29)$$

With loss-free media, k_2 and k_1 are real quantities, and consequently the only real roots of (29) are given by the intersection of the parabola, the left-hand side of (29), with the curve $\theta^2 \tan^2 \theta$, where $\theta = hd$. Evidently, there are only a few such roots in Fig. 5 (θ_0, θ_1 and θ_2) and these belong to the right Riemannian leaf. All the remaining roots (a countable infinity) are on the wrong Riemannian branch. The number of roots in the right leaf are determined by the relative values of k_2 and k_1 as well as the magnitude of d : the larger k_2 and d and the smaller k_1 , the larger is the number of these roots. With each and every one of these roots (h_0, h_1, h_2 , etc.) there is associated a pole ($\gamma_0, \gamma_1, \gamma_2$, etc.) in the plane of γ . All these poles are imaginary and are located on the imaginary axis between the points jk_1 and jk_2 (Fig. 6).

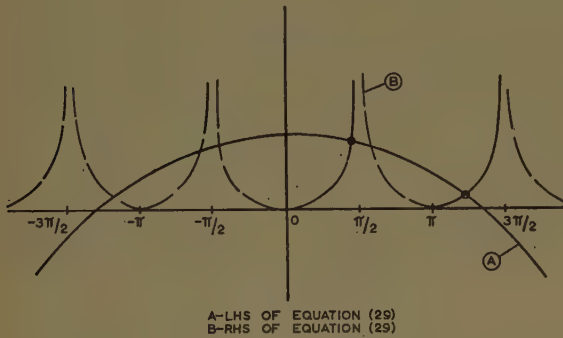


Fig. 5—Solution of (29).

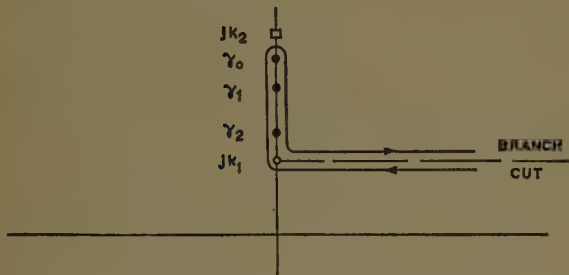


Fig. 6—The path of integration in the plane of γ , case $k_2 > k_1$.

Evidently, the path of integration can be distorted to that shown in Fig. 6 and consequently, it is permissible to interpret the entire field as made up from a number of pole-waves (due to the residues at the poles) and a branch-cut-wave which results from the integration around the branch-cut as indicated in Fig. 6.

The integrand of (26) may be put into the form

$$Q = \frac{\cos hx}{jh} \frac{\cos h(d-t) + Z \frac{u}{h} \sin h(d-t)}{\sin hd - Z \frac{u}{h} \cos hd} e^{-\gamma z} \quad (30)$$

with a branch point at $\gamma_B = jk_1$. To obtain the branch-cut-wave, we integrate (3) around the branch-cut, for which purpose we expand it in ascending powers of

(Zu/h) . This leads to an asymptotic series for the field H of the form

$$\psi_B = H_B \sim g_1 z^{-3/2} \Gamma\left(\frac{3}{2}\right) + g_2 z^{-5/2} \Gamma\left(\frac{5}{2}\right) + \dots \quad (31)$$

The first term of this expansion leads to

$$\psi_B = H_B \sim \frac{K}{\sqrt{2\pi}} \frac{\cos hx}{h} \frac{\cos h(d-t)}{\sin hd} [\cot hd + \tan h(d-t)] e^{j3\pi/4} \sqrt{\frac{k_1}{k_2^2 - k_1^2}} e^{-jk_1 z} \left(\frac{k_2}{k_1}\right)^2 z^{-3/2} \quad (32)$$

that is, the energy density of the wave is proportional to z^{-3} and the wave possesses a phase propagation coefficient equal to k_1 the intrinsic coefficient of the outside medium.

The pole-waves are evaluated from the knowledge of the residue at the poles. Thus, the n th pole is characterized by h_n , the n th root of (29). With such a pole, γ_n , there is associated a residue which gives for the intensity of the waves

$$H_n = jK \frac{\cos h_n x}{h_n d \gamma_n} \frac{h_n \cos h_n(d-t) - Zu \sin h_n(d-t)}{h_n \cos h_n d + Zu \sin h_n d} e^{-\gamma_n z} \quad (33)$$

The total field ψ can therefore be regarded as composed of the branch-cut-wave ψ_B and as many pole waves ψ_n as there are poles; real solutions to (29). We have therefore

$$\psi = \sum_{i=0}^n \psi_i + \psi_B \quad (34)$$

where each ψ_i is of the form $\psi_i = A_i e^{-\gamma_i z}$ with γ_i an imaginary number.

As mentioned above, such a procedure is permissible so long as none of the poles is in the immediate vicinity of the branch point, because under these circumstances the expansions employed do not converge.

Such a case would arise, for example, if we wished to examine the solution for the entire field as k_2 changes gradually from a very large value through all values down to the value $k_2 = k_1$.

The examination of the integrand shows that for $k_2 \gg k_1$, there are a large number of points clustered close to the point jk_2 with some poles distributed between jk_2 and jk_1 and an infinity of poles on the wrong leaf of the Riemannian surface. The poles are arranged in the order of their magnitude with the pole corresponding to the smallest value of h closest to jk_2 and the pole corresponding to the largest value of h closest to jk_1 . As k_2 decreases, all poles move towards the branch point jk_1 and the poles move one by one (through the branch point) to the wrong leaf of the Riemannian surface and, therefore, no longer constitute the essential part of the solution. Eventually, for sufficiently small values of k_2 in relation to k_1 and d , there is only one pole left and the

nearer to the branch point (jk_1) that this is situated, the smaller the difference $k_2 - k_1$. Eventually for $k_2 = k_1$, even this pole is captured by the branch point. In the other extreme for $k_2 \rightarrow \infty$ (or $k_1 \rightarrow 0$, which comes to the same thing), all poles collapse onto the one point jk_2 , which then becomes an essential singularity: this case, however, is of no practical significance.

The Existence of Pole Waves: In order to study the passage of a pole through the branch point, the field may be split into pole-waves as contributed by distant poles, and a wave which is a combined effect due to the pole nearest the branch point and the branch-cut itself (Fig. 7).

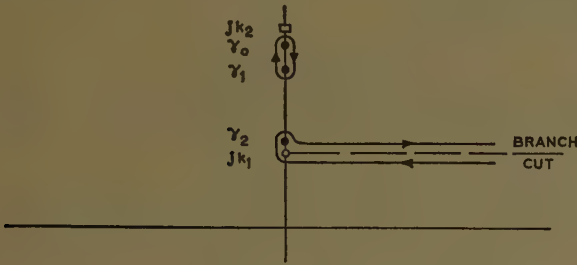


Fig. 7—A permissible modification of the path of integration shown in Fig. 6.

For the pole in the branch-cut proximity, we put

$$\gamma = jk_1 + v \quad (35)$$

where

$$|v| \ll k_1.$$

With this substitution in (30), expanding the trigonometric function about the point $\gamma = jk_1$, Q can be developed in a series and integrated.

For the n th pole, for which we put

$$h_0 = \sqrt{k_2^2 - k_1^2} = n\pi/d \quad (36)$$

and

$$h_n = h_0 + (\delta h)_n$$

the result is

$$H_n \simeq K e^{-j(k_1 + \delta^2)z} \cos h_0 x \cdot \cos h_0(d - t) \cdot \frac{h_0 d \delta h}{k_1 Z^2} \left(\frac{1}{\sqrt{\pi}} \operatorname{erfc}(jb\sqrt{z}) \right). \quad (37)$$

The subsequent terms of the series contain multipliers with negative powers of z and can, here, be neglected.

In expression (37), b is given by

$$b = \frac{h_0 d}{Z \sqrt{-2jk_1}} \delta h \quad (38)$$

and

$$\operatorname{erfc}(x) = \int_x^\infty e^{-t^2} dt. \quad (39)$$

Let (R_n) be the residue at the n th pole, then (37) will be recognized as

$$H_n = K(R_n) e^{-b^2 z} \operatorname{erfc}(jb\sqrt{z}) / \sqrt{\pi} = K(R_n) \Lambda_0 \quad (40)$$

which is the pole-wave with an amplitude factor $e^{-b^2 z} \operatorname{erfc}(jb\sqrt{z}) / \sqrt{\pi} = \Lambda_0$.

Because of the relations (36), for values of k_2 sufficiently large so that $\sqrt{k_2^2 - k_1^2} > n\pi/d$, δh is positive (the pole is in the right leaf of the Riemannian plane), and consequently for large values of z

$$\frac{1}{\sqrt{\pi}} \operatorname{erfc}(jb\sqrt{z}) \sim \left\{ 1 - \frac{1}{\sqrt{\pi}} \frac{e^{-b^2 z}}{2b\sqrt{z}} \right\} \quad (41)$$

that is, for sufficiently large values of z , the contribution from the pole becomes a pole-wave as defined before. Whereas for k_2 values such that $\sqrt{k_2^2 - k_1^2} < n\pi/d$, δh becomes negative and consequently

$$\frac{1}{\sqrt{\pi}} \operatorname{erfc}(jb\sqrt{z}) \sim \frac{1}{\sqrt{\pi}} \frac{e^{b^2 z}}{2b\sqrt{z}} \quad (42)$$

that is, the contribution from the pole vanishes.¹²

The magnitude of the multiplier Λ is shown plotted in Fig. 8 as a function of $\sqrt{k_2 - k_1} = \delta k$.

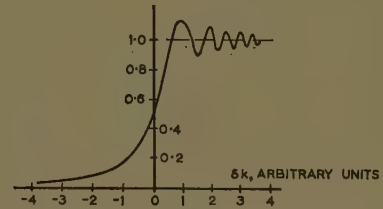


Fig. 8—The magnitude of the multiplier Λ as a function of δk .

In the present problem, the pole-waves are commonly known as surface waves, and in the light of the analysis given, it transpires that the question of existence cannot be answered uniquely except for the case of $z \rightarrow \infty$, in which case for positive δh (pole above the branch-cut) the pole-wave exists [(41) applies], whereas for negative δh , the pole-wave does not exist [(42) applies]. For finite values of z , however, little meaning can be attached to the existence of a pole-wave.

The branch-cut wave exists always and is representable asymptotically by a series in inverse fractional powers of z [(31)].

Propagation in "Lossy Media": With physical waveguides, due to the losses, k_1 and k_2 will be complex, although if medium 1 is air, then k_1 may be taken equal to $k_0 (= \omega \sqrt{\mu_0 \epsilon_0})$ a real number. In the latter case, the poles in the plane of γ are all complex [roots of (28)] as

¹² Asymptotically (42) vanishes as $z^{-1/2}$, but there are other terms in the expansion for H and these cancel (42) to the order of $z^{-1/2}$. The remaining terms of the series start with $z^{-3/2}$ and the series becomes identical with (31) and (32).

indicated in Fig. 9, but the integration (26) can still be carried out around the path shown in Fig. 9. The solution to the problem is again of the type (34) but now the individual pole-waves ψ_i have complex amplitude (a_i) factors [(33)] and their propagation coefficients γ_i (the poles) are also complex. Nevertheless, provided that none of the poles are in the immediate vicinity of the branch-cut, (31) to (34) apply.

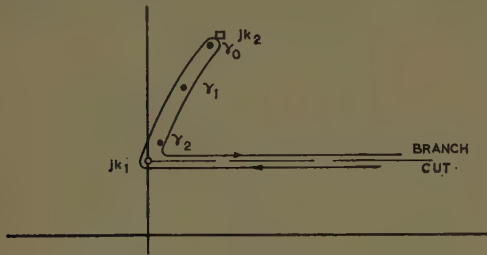


Fig. 9—Path of integration in the case of a waveguide with losses.

More precisely, however, the contribution of each of the poles to the total field is given by

$$H_n = K(R_n)\Lambda \quad (43)$$

where

$$\Lambda = e^{-b^2 z} \operatorname{erfc}(jb\sqrt{z})/\sqrt{\pi} \quad \text{if } \operatorname{Re}(b^2) \leq 0 \quad (44)$$

and

$$= e^{-b^2 z} [1 - \operatorname{erfc}(b\sqrt{z})/\sqrt{\pi}] \quad \text{if } \operatorname{Re}(b^2) > 0. \quad (45)$$

The quantity R_n is as before the residue at the n th pole.

We observe that with lossy waveguides (as with lossless waveguides), as k_2 decreases, the poles one after another pass over the branch-cut to the wrong Riemannian leaf and in consequence of the factor Λ disappear from the solution. However, with lossless waveguides, the poles pass to the wrong Riemannian leaf through the branch point itself, whereas with lossy waveguides the poles cross the branch-cut to the right of the branch point; the nearer the branch point, the smaller the losses. The pole-waves are also attenuated waves and therefore disappear from the asymptotic solution, leaving the branch-cut wave as the only significant part of the field.

If, in addition, medium 1 is also lossy, then the equations as given above are still applicable, although now k_1 is complex. As a result, the branch-cut wave (since it contains the factor $e^{-jk_1 z}$) becomes an attenuated wave. Consequently, now the total field is given by (34) with

$$\psi_i = A_i e^{-\gamma_i z} = K(R_i)\Lambda_i e^{-\gamma_i z} \quad (46)$$

$$\psi_B = e^{-jk_1 z} \sum_{n=1}^{\infty} g_n z^{-(2n+1)/2} \Gamma \frac{2n+1}{2}. \quad (47)$$

Whether the branch-cut wave or the pole wave now dominates in the asymptotic solution is entirely a matter of relative losses in media 1 and 2. That is, if the losses in medium 1 are sufficiently high, then pole-waves

will dominate for $z \rightarrow \infty$. Otherwise, it is the branch-cut wave that describes the field at infinity.

Outside Medium 1 Denser Than the Inside Medium 2

If $|k_1| > |k_2|$, then (28) again possesses a discrete infinity of solutions, but it is important to observe that if k_1 and k_2 are real, then none of the poles are in the right Riemannian leaf. Consequently, there are no pole-waves, and the only solution to the field is the branch-cut wave as given by (32).

Suppose, now, that k_2 is real (say medium 2 is air, $k_2 = k_0$), but that k_1 is complex and large. Under these conditions, (28) can be solved simply by successive approximations, and to the first order of quantities, we have for the n th root

$$h_n = h_0 + \delta h$$

$$h_0 = \frac{n\pi}{d}, \quad \delta h = \frac{j}{d} \left(\frac{k_2}{h_0} \right) \left(\frac{k_2}{k_1} \right)$$

$$\gamma_n = \sqrt{h_0^2 - k_2^2} + 2h_0\delta h$$

$$\approx \gamma_0 + \frac{h_0}{\gamma_0} \delta h$$

$$\text{if } \gamma_0 \neq 0$$

$$\gamma_0 = \sqrt{h_0^2 - k_2^2}. \quad (48)$$

The validity of the above formulas is subject to $|k_2| \ll |k_1|$.

The distribution of poles in relation to the branch-cut is shown in Fig. 10, and it will be observed that there is only a finite number of poles in the right leaf of the Riemannian plane. The actual number of poles on the right branch is determined by the magnitude of the imaginary part of k_1 . In particular, if

$$k_1 = k(1 - j\delta) \quad (49)$$

then the number of poles on the right branch of γ are given by the inequality

$$k\delta > \left(\frac{k_2}{k} \right)^2 \sqrt{\frac{k^2 + h_0^2}{kd}} \quad (50)$$

where $h_0 = n\pi$, or

$$n < \frac{kd}{\pi} \sqrt{\delta^2 (kd)^2 \left(\frac{k}{k_2} \right)^4 - 1}. \quad (51)$$

It will be observed that even for poor dielectrics, there are only very few poles present on the right branch. With conductors, however, k is very large ($k_1 = (1+j)\sqrt{\pi f \mu \sigma}$), and δ is not small. As a result, the number of poles on the correct branch is an extremely large number, yet finite. As $I_m(k_1)$ decreases, more and more poles pass on to the wrong Riemannian plane until eventually, for $I_m(k_1) = 0$, all poles are on the wrong branch.

On the other hand, if $ph(k_1)$ is kept constant but the magnitude of k_1 is decreased, then all poles move away from the real and imaginary axes towards the branch-cut, until eventually they disappear.

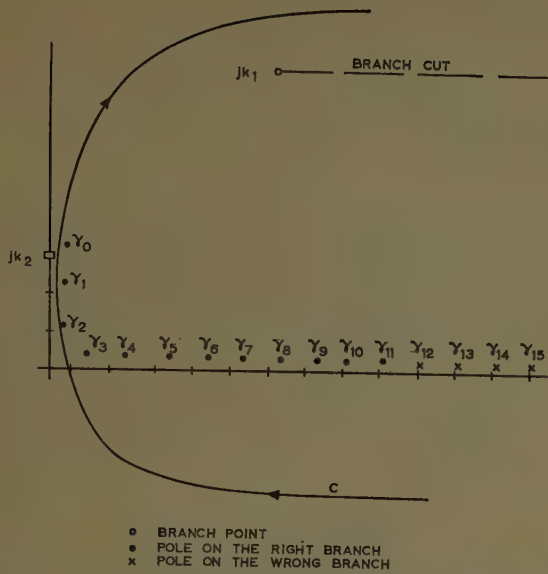


Fig. 10—Path of integration in the plane of γ appropriate to a metallic waveguide.

The solution to the field problem is again given, as before, by (34), with ψ_B given by (32), and ψ_i given by

$$\psi_i = jK \frac{\cos h_n(d - i)}{h_n d \gamma_n} \cos h_n x \cdot e^{-\gamma_n z}. \quad (52)$$

The remarks given in the previous section are also applicable in this case.

If, in addition, the medium filling the waveguide is lossy, then all the formulas given are applicable, provided that for k_2 the appropriate value is substituted. The position of all the poles is thereby altered.

DISCUSSION AND CONCLUSIONS

Although it is customary to talk about closed and open guiding systems,¹³ it is instructive to observe that such a classification is, strictly speaking, only valid in the limit of perfectly conducting surfaces. A perfectly conducting metal tube forms a closed waveguide (but it is not a physical reality). Such a waveguide (Fig. 2) is characterized at any particular frequency by a finite number of nonattenuating propagation modes and on infinity of evanescent modes. The latter can be ignored at large distances from the source.

With plain metallic tubes due to the finite (although large) conductivity in addition to the infinity of poles, there is also a branch point in plane of γ . This branch point is located at a very large distance from the origin on a line passing through the origin at 45° to the real axes, and there is only a finite number of poles in the right Riemannian leaf with the remaining poles located on the wrong leaf.

As the conductivity decreases, the branch point moves towards the origin, and at the same time, more

and more poles pass onto the wrong Riemannian sheet. Simultaneously, all the poles move slowly away from the real and imaginary axes, towards the branch point. It is important to observe, however, that the number of poles on the right branch are determined by the magnitude of the imaginary part of k_1 (30), whereas, the proximity of poles to the axes is related to the magnitude of k_1 , and if k_1 were a very large real number, the poles would lie on the co-ordinate axes, but none on the right branch.

For these reasons, provided that $I_m(k_1)$ is a large quantity, the treatment of tubular metallic waveguides as "closed" waveguides whose walls exhibit small surface impedance is an extremely good approximation. Consequently, in such waveguides, the resolution of fields into pole waves (Fig. 3) is admissible. Yet, it must be remembered that such treatment neglects the branch-cut wave and ignores the fact that mathematically, the expansion in terms of an infinity of pole-waves is a divergent expansion, and therefore, really not permissible.

It has been established that a branch-cut is always associated with all physical waveguides, and the general form of such a wave is

$$\psi_B = g_1 z^{-3/2} \Gamma\left(\frac{3}{2}\right) + g_2 z^{-5/2} \Gamma\left(\frac{5}{2}\right) + \dots \quad (53)$$

where the coefficients g contain an exponential factor $\exp(-jk_1 z)$.

In the far field, therefore, only the first term of the above series needs to be considered, and if the outside medium is loss-free, then the field intensities are proportional to $z^{-3/2}$.

All pole-waves have the general form

$$\psi_i = a_i \cdot f_i(x) \cdot \Lambda_i \exp(-\gamma_i z) \quad (54)$$

where a_i is the residue at the pole multiplied by K , and $f(x)$ has the form

$$f(x) = \cos h_n x \quad \text{if } x \leq d$$

and

$$f(x) = e^{-u x} \quad \text{if } x \geq d. \quad (55)$$

The quantity Λ_i is given by

$$\Lambda_i = e^{-b^2 z} \operatorname{erfc}(jb\sqrt{z})/\sqrt{\pi} \quad (56)$$

and as such is a function of z and the distance between the pole and the branch-cut. In the far field, Λ_i may be taken equal to unity if the pole is above the branch-cut (right Riemannian surface) and zero if the pole is below the branch-cut (wrong Riemannian surface).

With waves as defined above, there can never be any dispute about the existence of a branch-cut wave since this exists with all physical waveguides and the position of the branch point is determined by the properties of the outside medium (k_1). Similarly, all pole waves with nonzero residue exist, although the proportion of

¹³ A coaxial or a tubular metallic waveguide are examples of closed waveguides, whereas twin-wire transmission line and single-wire transmission line are examples of open structures.

the wave in the total field depends not only on the value of the residue but also on the quantity Λ [Fig. 8 and (56)]. Thus, if the pole is above the branch-cut, the wave exists with an amplitude almost equal to the residue, but if the pole is below the branch-cut (wrong branch), then the amplitude is extremely small; for large enough values of z , the wave may be almost said to be absent from the field investigated.

At this stage, some consideration should be given to the relative importance of the contribution to the total field from the individual singularities of the basic integral (26). In general, apart from other factors, the further is a singularity positioned to the right of the origin (Fig. 10), the less its relative importance. Thus, in the far field, it is permissible to retain a pole γ_i (for instance) and contributions from all singularities positioned to the left of γ_i , and if $\text{Re}(jk_1)$ happens to be greater than $\text{re}(\gamma_i)$, then to the accuracy considered, the branch-cut wave may be neglected. Such a reasoning justifies the surface impedance approach¹⁰ to "closed" waveguides and is the reason why the problem of existence of guided waves in connection with tubular metallic waveguides has never arisen.

On the other hand, it is useless to argue that a pole placed on the wrong Riemannian sheet could be brought on to the right sheet simply by deforming the branch cut B_1 to B_2 and regarding the field as composed of the pole-wave due to γ_{i+1} and a branch-cut wave due to B_2 , and then stating that γ_{i+1} exists as if it were on the right branch. Such an argument amounts to re-definition of branch-cut waves as can be seen from Fig. 11; B_2 is equivalent to B_3 and therefore

$$\int_{B_2} = \int_{B_3} = \int_{B_1} - \psi_{i+1}. \quad (56)$$

In this way, all that has been achieved is a re-definition of the branch-cut wave. By similar reasoning, the waves associated with γ_{12} , γ_{13} , etc., in Fig. 10 could be brought to the right Riemannian sheet and given significance, despite the fact that these waves (growing exponentially to infinity in the positive x direction) cannot exist physically. Mathematically, there is nothing wrong with such an argument but physically it is a useless argument.

Let us now examine Sommerfeld's problem in a little more detail. In this case, a solution is sought to the field problem (Figs. 6, 7 and 9) for the case when all the poles except γ_0 have passed on to the wrong Riemannian surface. Suppose that k_1 is real and $k_2 = k_1 + q$ where q is a small number. It can then be shown that $h = \sqrt{2k_1q}$, $u = -j\sqrt{j2k_1M}q^2$ and $=jk_1(1 + Mq^2)$ where $M = 2d^2 + k_1^{-2}$. Consequently, if $|q|$ is kept constant and $ph(q)$ is made to vary between 0 through 90° to 180°, the pole traverses a small arc around the branch point (on to the wrong

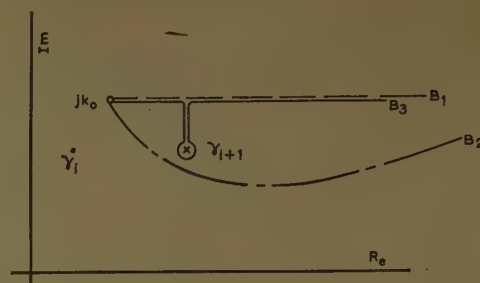


Fig. 11—Equivalent branch-cuts.

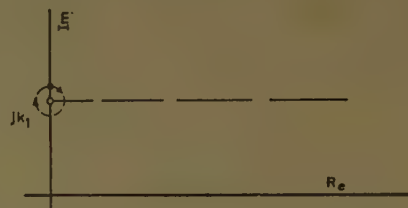


Fig. 12—The passage of the pole over the branch-cut.

Riemannian surface) as indicated in Fig. 12. Evidently, $ph(q) = 0$ corresponds to a loss-free reactive surface, the point $ph(q^2) = 90^\circ$ corresponds to a surface having surface impedance such that $ph(Z_s) = 45^\circ$ (the case of resistive earth in Sommerfeld's problem), and $ph(q^2) = 180^\circ$ corresponds to $ph(Z_s) = 0$ (the case of dielectric earth). Consequently the pole-wave (which can here be identified with the Zenneck wave) does not exist in the far field of a dipole, in the sense that its amplitude is extremely small.

A NOTE ON NOMENCLATURE

Waves associated with poles on the right branch are termed real waves and those associated with the wrong branch the virtual waves. In this way, real surface waves are slow waves and virtual surface waves are fast waves (e.g., Zenneck wave) but both are true guided waves. Waves associated with tubular metallic waveguides lose power laterally and can be termed "leaky waves." These can be subdivided into evanescent and propagating modes, and provided the losses are not too large, the subdivision has practical significance. With the exception of the ψ_0 wave, all such waves are fast waves, and they are real waves if the poles are on the right branch, and virtual waves if they are on the wrong branch.

Pole-waves are defined by (54) and quasi-modes by (54) with $\Lambda = 1$.

ACKNOWLEDGMENT

The author wishes to acknowledge the Standard Telecommunication Laboratories for permission to publish this paper.

Guided Waves on Sinusoidally-Modulated Reactance Surfaces

A. A. OLINER[†] AND A. HESSEL[†]

Summary—A rigorous solution is derived for the propagation characteristics and field distributions of waves guided by a plane surface which possesses a surface reactance modulated sinusoidally in the propagation direction. The explicit field amplitudes and the determinantal equation for the propagation wavenumber are expressed in a continued fraction form which is rapidly convergent for all values of modulation. Numerical results are obtained for both surface wave (modal) and leaky wave (nonmodal) solutions. The relevance of these studies to high-gain modulated surface-wave antennas is discussed.

I. INTRODUCTION

THE present study of guided waves on sinusoidally-modulated reactive surfaces was motivated by recent developments in high-gain end-fire antennas of the surface wave type. A customary method for improving the gain of surface-wave antennas has been to design them according to the Hansen-Woodyard condition. For high gains, however, the surface wave becomes necessarily very loosely bound to the guiding surface, and the performance becomes sensitive to irregularities in the structure. A break-through with regard to this difficulty was offered by Simon^{1,2} several years ago; he showed that gains as high as those predicted by the Hansen-Woodyard condition for a comparable antenna length, together with a tightly-bound surface wave, were possible when the surface wave structure was modulated along its length. Since then, a variety of modulated surface-wave antennas have been designed, but these designs are largely empirical, being a combination of sinusoidal modulations, tapers, steps, and in-between variations. The theoretical understanding of the control of these effects is still limited, and the state of the art is hampered by this lack of knowledge.

Thomas and Zucker³ have examined the effect of modulating the phase velocity of a surface wave by employing a spectral representation of the modulation. They show that with either a phase or amplitude modulation, a radiated beam can be produced which is tilted at an arbitrary angle with the surface, and may be made end-fire if desired. They did not, however, as they acknowledge, study the surface parameters required to yield the necessary phase or amplitude distribution. A

start in this direction was made by Pease,⁴ who considered a dielectric slab, the index of refraction of which was modulated very gently in a sinusoidal fashion. The approximations employed in his analysis, however, which were made to correspond to slow variations, are sufficiently severe as to cast doubt on the practicality of his results. The present study, which is a continuation in this direction, is a rigorous treatment of sinusoidal variations in surface reactance.

The surface under study is a plane surface possessing a reactance that varies sinusoidally in the direction of propagation of the guided waves. If the period, or spacing, of the modulation is sufficiently small, the modulation serves to affect quantitatively the propagation wavenumber of the surface wave guided along it, and to introduce one or more stop bands. If this spacing is larger than some critical value, the modulated surface wave will give rise to one or more leaky waves which radiate away from the surface at some angle. Because of its relevance to the antenna problem, only the E (or TM) waves guided by the surface are considered here.

A solution to the problem is obtained by viewing the geometry in terms of modes associated with propagation perpendicular to the surface; these modes are discrete in character and are specified by the periodicity of the modulation. The nature of the impedance boundary condition at the surface is such that each of these modes couples only to itself and to the next higher and lower modes, in contrast to most diffracting surfaces which couple all modes to each other. As a result, the problem reduces to the solution of an infinite set of linear equations in which each equation involves only three modes. Explicit expressions for the field amplitudes of each of these modes (which are equivalent to the space harmonics) and the determinantal equation for the propagation wavenumber along the surface are then expressed in a continued fraction form which is rapidly convergent for all values of modulation. In addition, a very simple explicit expression for the propagation wavenumber is deduced by a perturbation procedure that is valid for small modulation values.

The physical mechanism for the leakage of energy away from the surface wave may be regarded in a fashion analogous to that for the leaky wave produced by a slitted rectangular waveguide, for example. In this latter example, the opening offered by the slit permits the radiation to escape. In the surface wave case, the periodic modulation produces an infinity of discrete

[†] Microwave Res. Inst., Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

¹ J. C. Simon and V. Biggi, "Un nouveau type d'arien et son application à la transmission de télévision à grande distance," *L'Onde Électrique*, no. 332; November, 1954.

² J. C. Simon and G. Weill, "Un nouveau type d'antenne à rayonnement longitudinal," *Ann. de Radioélectricité*, vol. VIII; July, 1953.

³ A. S. Thomas and F. J. Zucker, "Radiation from Modulated Surface Wave Structures—I," 1957 IRE NATIONAL CONVENTION RECORD, Pt. 1, pp. 153-160.

⁴ R. L. Pease, "Radiation from modulated surface wave structures—II," 1957 IRE CONVENTION RECORD, Pt. 2, pp. 161-165.

modes of the type referred to above, all of which are below cutoff if the modulation period is sufficiently small. (An equivalent remark would be that all of the space harmonics are slow waves.) If the modulation period is increased, one or more of these modes may become propagating; power is thus radiated away from the surface wave itself via the mechanism of a higher-order mode introduced by the periodicity.

The characteristics of the surface wave and leaky wave solutions are discussed in Section III. Band structure curves are presented for the surface-wave solutions; of interest is the appearance of stop bands and the involved behavior at the band edges. The relative amplitudes of the various space harmonics are also presented. The discussion on the leaky wave solutions includes the variation of attenuation and phase constants with modulation spacing, or alternatively, with the angle of the radiated wave, and the relative amplitudes of the various radiated beams and the surface wave. The relation to the end-fire antenna problem is also touched upon.

II. FORMAL SOLUTION

Rigorous solutions are obtained below for the propagation characteristics and field distribution of waves guided by a plane surface which possesses a surface reactance sinusoidally modulated in the propagation direction. The reactive plane surface, which is chosen as the yz plane in Fig. 1, exhibits a surface reactance of the form

$$X(z) = X_s \left[1 + M \cos \left(\frac{2\pi}{a} z \right) \right], \quad (1)$$

where X_s is the constant value of surface reactance about which the modulation is made, M and a are the amplitude and period, respectively, of the modulation, and an $\exp(j\omega t)$ time dependence is chosen. The modulation amplitude M is restricted to the range $M \leq 1$. The impedance boundary condition at the plane $x=0$ is, therefore,

$$-E_t(0, y, z) = jX(z)H_t(0, y, z) \times \mathbf{x}_0, \quad (2)$$

where E_t and H_t refer to the transverse (to x) electric and magnetic fields, respectively, and \mathbf{x}_0 is the unit vector in the x direction.



Fig. 1—Modulated plane reactive surface.

The waves guided by the surface are taken to propagate in the z direction, and, for simplicity, no variation is assumed present in either the geometry or the fields in the y direction. An E and H (or TM and TE) modal decomposition is always possible in the x direction for this geometry, but, because of the lack of dependence on

y , these modes are identical with E and H modes viewed in the z direction. In the present paper, only the E mode solutions are considered because of their applicability to surface-wave antennas.

The total field is viewed in terms of modes defined with respect to propagation in the transverse, or x , direction. The corresponding vector mode functions \mathbf{h}_n and \mathbf{e}_n , which form an orthonormal set, are⁵

$$\begin{aligned} \mathbf{h}_n(z) &= \mathbf{e}_n(z) \times \mathbf{x}_0 \\ &= \mathbf{y}_0 \frac{1}{\sqrt{2\pi}} \exp \left[-j \left(\kappa + \frac{2n\pi}{a} \right) z \right], \end{aligned} \quad (3)$$

where κ is the propagation wavenumber along the surface (in the z direction), and is related to the transverse wavenumber k_{tn} by

$$\begin{aligned} k_{tn} &= \sqrt{k^2 - \left(\kappa + \frac{2n\pi}{a} \right)^2}, \\ n &= 0, \pm 1, \pm 2, \dots, \end{aligned} \quad (4)$$

k being the free-space wavenumber ($= 2\pi/\lambda$). With this notation, the total magnetic field, for example, is written

$$H_y(x, z) = \sum_{n=-\infty}^{\infty} I_n(x) h_n(z) \quad (5)$$

or

$$H_y(x, z) = e^{-ikx} \sum_{n=-\infty}^{\infty} \frac{I_n(0)}{\sqrt{2\pi}} e^{-j(2n\pi/a)z} e^{-jk_{tn}z}, \quad (6)$$

which could alternatively have been obtained via Floquet's theorem. The transverse mode viewpoint is chosen here because it permits a clearer physical interpretation of the solution.

The mode voltage V_n of the n th transverse mode is

$$V_n(x) = \int_0^a E_t(x, z) \cdot \mathbf{e}_n^*(z) dz, \quad (7)$$

which, in view of the boundary condition (2), becomes at $x=0$:

$$V_n(0) = j \int_0^a \left[1 + M \cos \frac{2\pi}{a} z \right] X_s H_y(0, z) \mathbf{e}_n^*(z) dz. \quad (8)$$

When the mode function \mathbf{e}_n from (3) is inserted into (8), and the relation for the transverse mode current I_n is recognized to be

$$I_n(x) = \int_0^a H_t(x, z) \cdot \mathbf{h}_n^*(z) dz, \quad (9)$$

expression (8) becomes

$$V_n = jX_s I_n + jX_s \frac{M}{2} (I_{n+1} + I_{n-1}). \quad (10)$$

⁵ N. Marcuvitz, "Waveguide Handbook," Rad. Lab. Series, McGraw-Hill Book Co., Inc., New York, N. Y., vol. 10, pp. 88, 89; 1951.

Relation (10) indicates that the nature of the boundary condition at $x=0$ is such that it couples each transverse mode to its nearest neighbors, rather than to all of the other transverse modes as usually occurs with diffracting surfaces.

The determinantal equation for the propagation characteristics of the guided waves, whether of modal or nonmodal (leaky wave) nature, may be phrased as the condition for resonance in the transverse plane. This transverse resonance condition states that at any cross-section (to x) plane, the sum of the impedances looking in both directions away from this plane must equal zero. At the $x=0$ plane, the impedance for the n th transverse mode looking toward the plane is obtainable from (10), while that looking outward is simply

$$\frac{V_n}{I_n} = Z_n = \frac{k_{tn}}{\omega\epsilon} \quad (11)$$

Combining (11) with (10), one finds for the transverse resonance condition

$$I_{n+1} + D_n I_n + I_{n-1} = 0, \quad n = 0, \pm 1, \pm 2, \dots, \quad (12)$$

where

$$D_n = \frac{2}{M} \left[1 - j \frac{k_{tn}}{X_s \omega \epsilon} \right]. \quad (13)$$

The recurrence relation (12) may be viewed as an infinite set of linear homogeneous equations for the infinite number of unknown modal currents I_n . This set of equations possesses a nontrivial solution if the infinite determinant of the set vanishes. This condition may be rephrased in the following manner. Consider the semi-infinite set of these equations which begins at any finite value of n and contains those equations with lower values of n . From the first two equations of this set, one can write

$$\frac{I_n}{I_{n+1}} = - \frac{1}{D_n - \frac{1}{D_{n-1} + \frac{I_{n-2}}{I_{n-1}}}}.$$

Continuing in this fashion for the remainder of the semi-infinite set, one finds the continued fraction solution

$$\frac{I_n}{I_{n+1}} = - \frac{1}{D_n} - \frac{1}{D_{n-1}} - \frac{1}{D_{n-2}} - \dots \quad (14)$$

From the remaining semi-infinite set of equations one has, analogously,

$$\frac{I_{n+1}}{I_n} = - \frac{1}{D_{n+1}} - \frac{1}{D_{n+2}} - \frac{1}{D_{n+3}} - \dots \quad (15)$$

It can be shown⁶ that if $|D_n| > 2$ for $n > n_0$, the continued fractions (14) and (15) converge; from (13) we see that $|D_n| \sim n$ for large n due to the presence of k_{tn} in the numerator of D_n .

In order that the infinite set of equations (12) have a non-vanishing solution, (14) must equal the inverse of (15). Thus, employing (13) and (4), one obtains for an inductive surface the following completely equivalent equations, one for each $n (= 0, \pm 1, \pm 2, \dots)$:

$$1 - \frac{j}{X_s'} \sqrt{1 - \left[\frac{\kappa}{k} + \frac{2\pi n}{ka} \right]^2} = \frac{M^2}{4} \left\{ \frac{1}{1 - \frac{j}{X_s'} \sqrt{1 - \left[\frac{\kappa}{k} + \frac{2\pi(n-1)}{ka} \right]^2}} - \frac{M^2/4}{1 - \frac{j}{X_s'} \sqrt{1 - \left[\frac{\kappa}{k} + \frac{2\pi(n-2)}{ka} \right]^2}} - \dots + \frac{1}{1 - \frac{j}{X_s'} \sqrt{1 - \left[\frac{\kappa}{k} + \frac{2\pi(n+1)}{ka} \right]^2}} - \frac{M^2/4}{1 - \frac{j}{X_s'} \sqrt{1 - \left[\frac{\kappa}{k} + \frac{2\pi(n+2)}{ka} \right]^2}} - \dots \right\}, \quad (16)$$

where $X_s' = X_s / \sqrt{\mu/\epsilon}$. Due to the presence of the additional factor of $M^2/4$ in each successive term, the continued fractions converge rapidly for all values of modulation. For calculational purposes, n is set equal to zero in (16), and the five terms shown are sufficient to yield the solution for the propagation wavenumber κ to a high degree of accuracy in almost all cases.

By means of (14) and (15), one may evaluate the amplitudes of each of the space harmonics relative to, say, I_0 . Since the ratio I_n/I_0 behaves asymptotically for large n as $1/n!$, the total field clearly converges. Thus, the solution presented above permits the calculation of the complete field distribution and the propagation characteristics to any desired degree of accuracy.

III. DISCUSSIONS OF RESULTS

The guided-wave solutions to (16) are of two types, the (modal) surface waves, for which κ is real, and the (nonmodal) leaky waves, for which κ is complex. When the modulation spacing a is small, the guided wave is a (trapped) surface wave; when this spacing exceeds a critical value, a leaky wave is created and radiation occurs. The propagation behavior for each of these two wave types is discussed below.

⁶ J. Meixner and F. W. Schäfke, "Mathiesche Funktionen und Sphäroidfunktionen," Springer Verlag, Berlin, Germany, pp. 89-93; 1954.

A. The (Trapped) Surface-Wave Solutions

For the surface-wave solution, κ is real and greater than k , corresponding to a slow wave, and all of the constituent space harmonics are also slow waves, with a field distribution that decays transversely to the guiding surface. Alternatively, all constituent transversely-directed modes are below cutoff in the x direction, so that k_{tn} is imaginary for all n , and the total field is confined to the neighborhood of the $x=0$ surface.

It can be shown that the allowed regions for the surface-wave solutions in a band structure plot is a succession of triangular areas with the tips of the triangles ending at $ka=\pi$, as shown in Fig. 2. Within each of these allowed areas, there also exists at least one stop band, the width of which is a function of the modulation depth. Typical band structure curves, for the case $X_s'=1$, are presented in Fig. 3 for three values of modulation amplitude M . The stop bands are seen to occur about $\kappa a = \pi \pm 2n\pi$; within the stop bands, the values of κ are complex. As expected, the width of the stop band is seen to be greater for larger values of M .

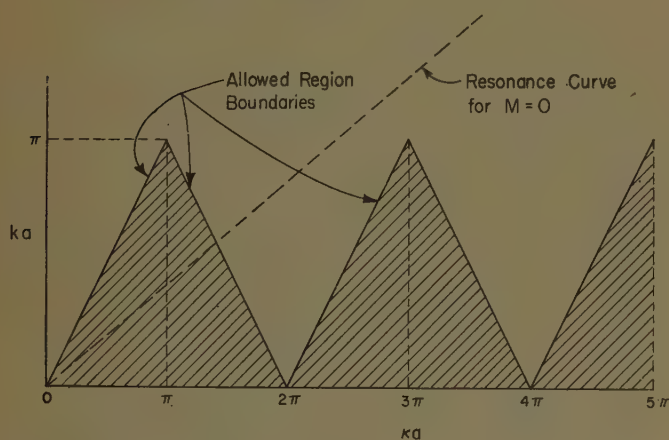


Fig. 2—Allowed regions for (trapped) surface-wave solutions.

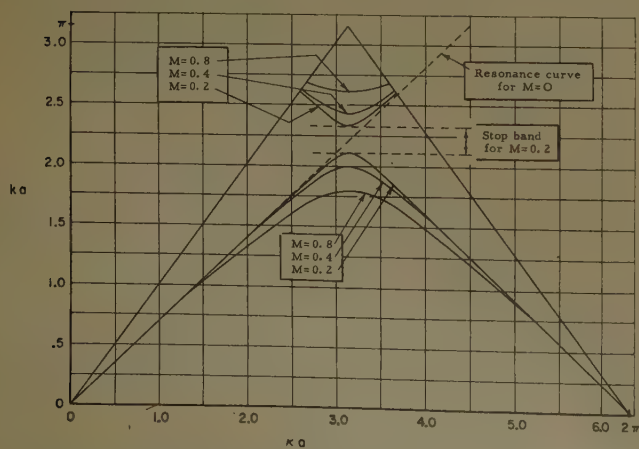


Fig. 3—Band structure curves for surface-wave solutions for various modulation depths, for $X_s'=1$.

The behavior at the band edges of the upper band is also of interest. Although not apparent in Fig. 2, due to the scale employed, the curves turn around at the band edges to become parallel to the allowed region boundaries.

The effect on the band structure curves of different average surface reactance values X_s' , for fixed M , is indicated in Fig. 4. It is seen that the curves follow the straight resonance lines obtained for $M=0$ for the corresponding values of X_s' . The middle curve, for $X_s'=1$ and $M=0.4$, is a duplicate of the middle curve of Fig. 3. The curve for $X_s'=0.2$ corresponds to a very loosely-bound wave, and, consequently, to a wave with a phase velocity very close to that of the velocity of light. It is noted that for this case the stop band is very small and is located near to the tip of the allowed region. Of particular interest is the $X_s'=5$ case, which corresponds to a very tightly-bound wave. It is seen that *more than one* stop band is obtained for this case. (For clarity, the repetition of the curves from one triangle to another, required for a complete band structure presentation, has been omitted.) Thus, as ka increases, the wave type changes from a purely trapped surface wave to a leaky wave and then back to a surface wave again for a band of ka . Such behavior, with associated multiple stop bands, is obtained when $X_s' > \sqrt{8}$. It is easily shown that n stop bands are present if $(2n-1)^2 - 1 < (X_s')^2 < (2n+1)^2 - 1$. Thus, three stop bands are actually present for $X_s'=5$, but since the third stop band occurs very near to the tip of the third triangle, it has been ignored in Fig. 4.

The relative amplitudes of the space harmonics constituting the surface wave are indicated graphically in Fig. 5 for the case $X_s'=1$, $M=0.4$, as a function of ka for the lower band. It is seen that to a good approximation only the fundamental and a single space harmonic ($n=-1$) need be considered. It is also noted that at the onset of the stop band, the amplitudes of the first backward-traveling space harmonic and the forward-traveling fundamental are equal, *i.e.*, $I_{-1}/I_0=1$, and that the amplitudes of the remaining forward- and backward-traveling space harmonics also become equal in pairs so as to establish a standing wave. This behavior is typical of that occurring in stop bands of periodic structures, and affords a check on the analysis.

As an illustration of the space harmonic content of the surface wave as a function of average surface reactance X_s' , the ratio I_{-1}/I_0 is plotted as a function of ka in Fig. 6 for three different values of X_s' , keeping M constant. It is seen that the amplitude of the $n=-1$ space harmonic increases as the wave becomes more tightly bound. The same behavior also results for the other space harmonics.

Since more than one stop band is found to exist for the case $X_s'=5$ (see Fig. 4), a plot of the relative amplitudes for this case serves to indicate the differences in

space harmonic content that occur in the upper and lower pass bands and in successive triangles. The relative amplitude curves for the first triangle are given in Fig. 7. It is seen that the curves for the lower pass band are similar in form to those appearing in Fig. 5 for a different value of X'_s . The curves in the upper band, however, have a different form from those in the lower band. The relative amplitudes for the second triangle (see the corresponding band structure curve in Fig. 4)

are presented in in Fig. 8. It is noted that the space harmonic content is quite different in the second triangle, and that more space harmonics are required to properly characterize the field.

Since a standing wave is established at the band edges of the stop band, certain necessary relations between the amplitudes must exist. It can readily be shown that these relations are: For the first triangle:

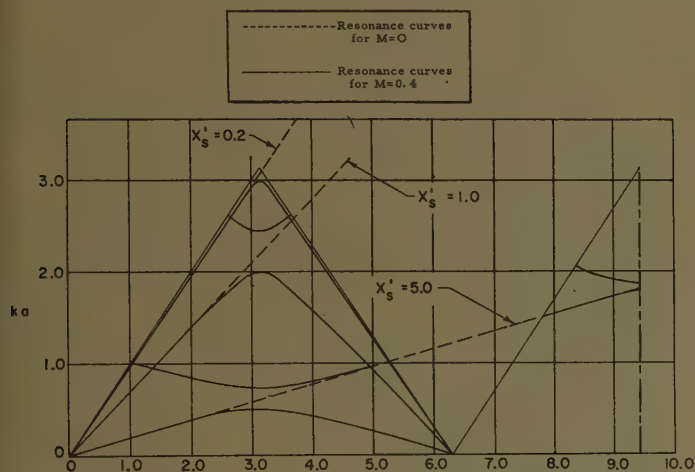


Fig. 4—Band structure curves for surface-wave solutions for various average surface reactances, for $M=0.4$.

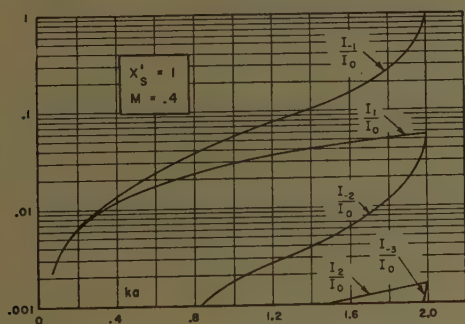


Fig. 5—Relative amplitudes of the space harmonics of the surface wave in the lower band.

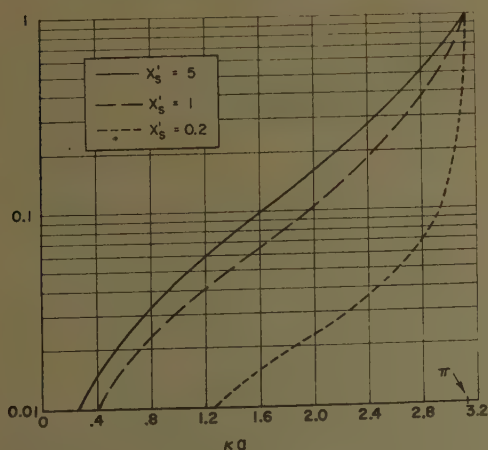


Fig. 6—Relative amplitudes of the $n = -1$ space harmonic for various surface reactances, for $M=0.4$.

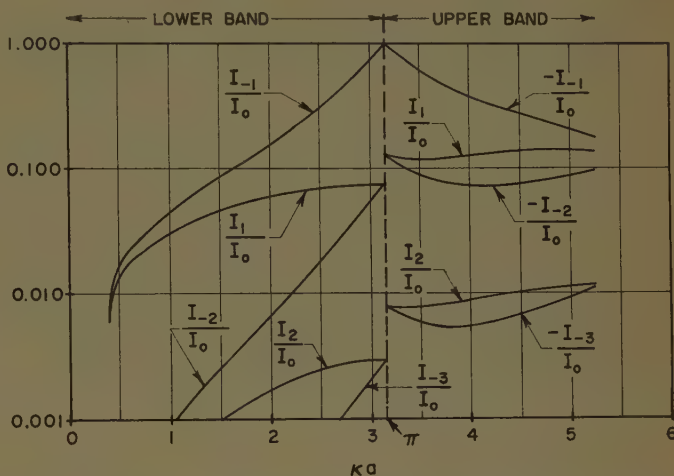


Fig. 7—Relative amplitudes of the space harmonics of the surface waves in the upper and lower bands in the first triangle for $X'_s=5$, $M=0.4$.

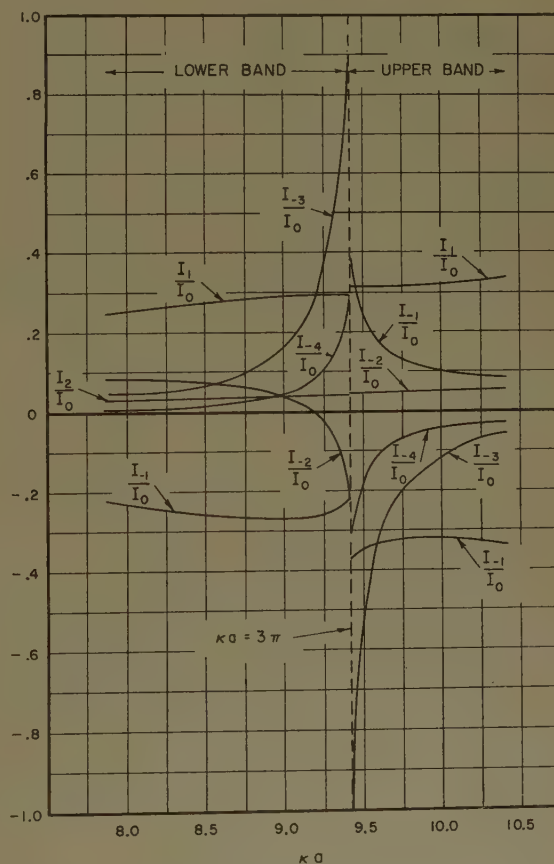


Fig. 8—Relative amplitudes of the space harmonics of the surface waves in the second triangle for $X'_s=5$, $M=0.4$.

$$I_{-1} = \pm I_0, \quad I_{-2} = \pm I_1, \quad \dots \quad I_{-n} = \pm I_{n-1}.$$

For the second triangle:

$$I_{-3} = \pm I_0, \quad I_{-2} = \pm I_1, \quad \dots \quad I_{-n} = \pm I_{n-3}.$$

The upper and lower signs refer to the lower and upper bands, respectively. The involved behavior of the curves in Fig. 8 is readily explainable in terms of these relations.

B. The Leaky Wave Solutions

When ka is raised sufficiently so that a surface wave solution is no longer possible, the $n = -1$ transversely-directed mode is raised above cutoff, corresponding to a *propagating* wave in the direction perpendicular to the surface, while the remainder of these modes are still substantially below cutoff in this direction. As a result, the value of κ from (16) becomes complex. As the value of ka is raised further, additional transversely-directed modes may become propagating. Each of these modes remains above cutoff for only a limited range of ka , however, so that only a small number of these modes may be propagating for any value of ka . This number is a function of the average surface reactance X_s .

If β_u represents the unperturbed value of κ , corresponding to $M=0$, the value of κ in the leaky wave region may be written as

$$\kappa = \beta_u + \Delta\beta - j\alpha. \quad (17)$$

When the transverse wavenumbers k_{tn} are examined consistent with form (17) for κ , one finds that those transversely-directed modes which are *above* cutoff correspond to *outgoing* waves in the transverse direction, with an exponential amplitude increase in this direction if the wave is directed forward and a decrease if it is directed backward. All of the other transversely-directed modes exhibit the exponential transverse decay associated with their cutoff nature, but possess a small phase variation corresponding to an incoming wave for the forward-traveling space harmonics and to an outgoing wave for the backward-traveling space harmonics.

Those space harmonics corresponding to $n \geq 0$ are always forward-traveling waves and always remain essentially below cutoff. A given space harmonic for which n is negative is a backward-traveling wave which is below cutoff (in the surface wave region) or essentially below cutoff (in the leaky wave region) for the smaller values of ka . If ka is raised sufficiently, the wave changes into an essentially propagating wave which first propagates in the *backward* direction. As ka is increased further, the radiating wave swings up, passes through broadside and approaches the forward end-fire position. As ka is raised still further, the wave becomes essentially below cutoff again and, for all higher ka values, remains a *forward-traveling* space harmonic.

It is convenient to employ a perturbation solution of (16) for the rapid evaluation of the complex κ value for small values of modulation. The perturbation is taken about the $M=0$ value to yield

$$\Delta\beta a - j\alpha a = -\frac{M^2}{4} \frac{kaX_s'^2}{\sqrt{1+X_s'^2}}.$$

$$\left[\frac{1}{1 - \frac{j}{X_s'} \sqrt{1 - \left[\sqrt{1 + X_s'^2} - \frac{2\pi}{ka} \right]^2}} + \frac{1}{1 - \frac{j}{X_s'} \sqrt{1 - \left[\sqrt{1 + X_s'^2} + \frac{2\pi}{ka} \right]^2}} \right]. \quad (18)$$

Expression (18) takes into account only the $n=0, 1$, and -1 transversely-directed modes (or space harmonics), and yields accurate results except within certain narrow ranges of ka .

A plot of the attenuation constant as a function of ka is presented in Fig. 9 for the case $X_s'=1$, $M=0.4$. To this order, the behavior is strongly associated with that for the $n=-1$ transversely-directed mode. Perturbation formula (18) is inadequate when the radiating beam corresponding to the $n=-1$ term points in the broadside direction, or in either the forward or backward end-fire directions. In these "transition" regions more space harmonics are required to describe the behavior accurately. For this reason, the curve of Fig. 9 avoids both end-fire regions; it should also be understood, however, that the curve is incorrect in the vicinity of broadside radiation, *i.e.*, around $ka=4.4$.

The modulation also serves to alter the guide wavelength slightly, and the change $\Delta\beta$ in the phase constant computed via relation (18) is plotted in Fig. 10 as a function of ka . It is seen that the change is such as to decrease β for the smaller ka values, but to increase it

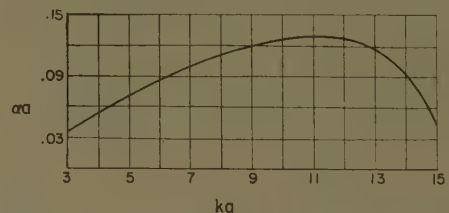


Fig. 9—Attenuation constant of leaky wave as a function of ka , computed to first order, for $X_s'=1$, $M=0.4$.

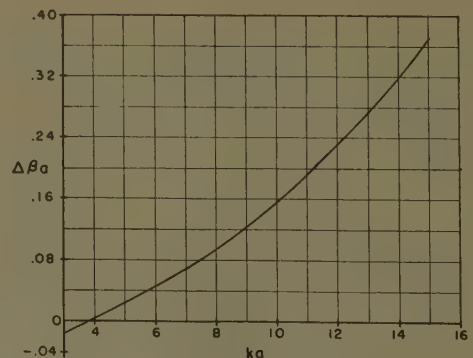


Fig. 10—Change in phase constant of leaky wave as a function of ka , computed to first order, for $X_s'=1$, $M=0.4$.

as ka is increased. The remarks regarding the inadequacy of (18) for the attenuation constant apply also to $\Delta\beta$.

The amplitude ratios of the space harmonics, or alternatively, of the transversely-directed modal constituents of the leaky wave, are now complex, but to first order, consistent with perturbation result (18), only those ratios involving radiating beams are complex. It is found that for all but certain narrow ranges of ka (corresponding to the above-mentioned "transition" regions), for which the amplitude of a particular space harmonic may become large, all field amplitudes are small except for I_0 , I_{-1} and I_1 , with I_0 dominant.

The magnitudes and phases of the relative amplitudes in the leaky wave region for the case $X'_s = 1$, $M = 0.4$ are presented in Figs. 11 and 12. With regard to the magnitudes plotted in Fig. 11, it is seen that only $n=1$ and $n=-1$ terms are significant relative to the fundamental ($n=0$) over the whole range of ka covered in the figure. The ratio $|I_{-2}/I_0|$ remains small except in the vicinity of $ka=4.4$, which is a "transition" region corresponding to broadside radiation for the $n=-1$ beam. To first order, this ratio becomes infinite; when more space harmonics are taken into account, the resulting more accurate calculation yields a finite maximum value. From Fig. 12, one notes that the ratios I_1/I_0 and I_2/I_0 are real. A π phase shift in I_{-2}/I_0 is seen to occur at the value of ka corresponding to the peak in the magnitude

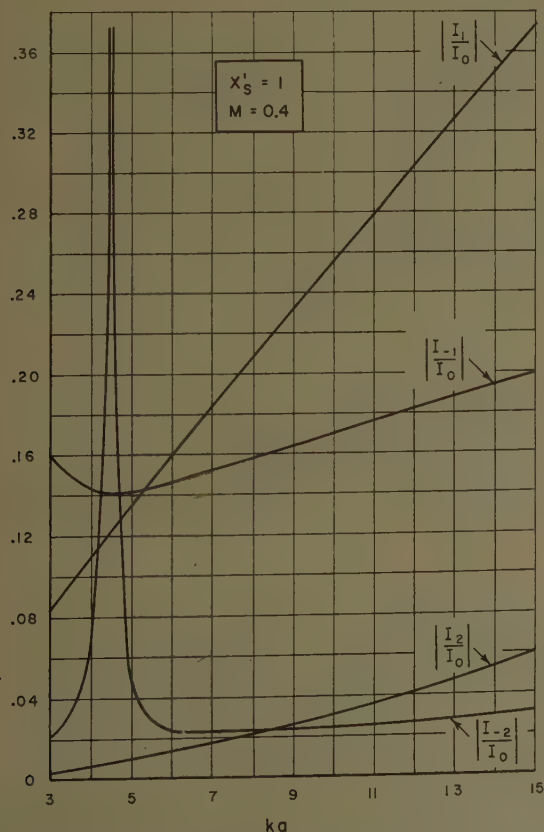


Fig. 11—Magnitude of the complex amplitudes of the space harmonics of the leaky wave as a function of ka , computed to first order, for $X'_s = 1$, $M = 0.4$.

of this ratio. Except for this jump, the values of phase are seen to be constant or slowly-varying.

A few brief remarks are made below on the application of these studies to the antenna problem. As mentioned above, the first radiating beam that appears as ka is raised is at first directed backwards, along the surface in the negative z direction, as is well-known from array considerations. As ka is raised further, this beam swings up through broadside and eventually reaches the forward end-fire position before vanishing. However, unless λ/λ_g is greater than three (corresponding to $X'_s > \sqrt{8}$), at least one additional beam will be present, directed at some angle, while the original beam is in the forward end-fire position. (If no additional beam is present, more than one stop band will occur and the appropriate band structure curves are similar to those shown in Fig. 4 for $X'_s = 5$.) The case for which two beams are present is illustrated in Fig. 13.

It is clear that, by proper choice of the average surface reactance value X'_s , one can eliminate the presence of unwanted additional beams. However, this requires a very tightly-bound surface wave, with attendant higher space-harmonic content. It is customary to employ a less tightly-bound wave, with the result that one or

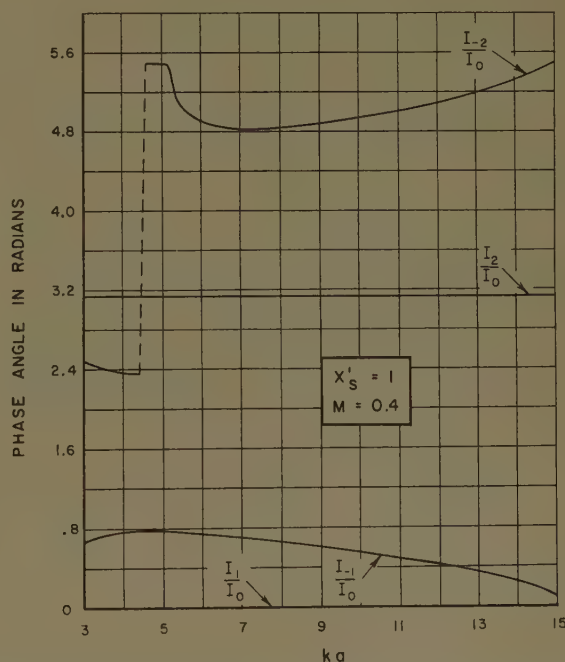


Fig. 12—Phase of the complex amplitudes of the space harmonics of the leaky wave as a function of ka , computed to first order, for $X'_s = 1$, $M = 0.4$.

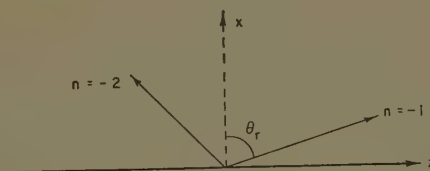


Fig. 13—Illustration of two radiating beams in the leaky wave region. Angle θ_r is the real part of complex angle θ , defined by $\cos \theta = k_z/k$.

more additional beams may be present. If a significant amount of power is carried by any of these additional beams, their influence on the desired radiation pattern will not be negligible. While in the presence of radiation loss power orthogonality no longer exists rigorously between the individual transversely-directed modes, the square of the amplitude ratios is nevertheless a reasonably good measure of the power radiated if the attenuation constant remains small.

The antenna design may require the radiated beam to appear at any desired angle, say at $\theta_r = 30^\circ$ in Fig. 13. For such an angle, away from the "transition" regions, the perturbation results are accurate. For the case $X_s' = 1$, $M = 0.4$, this angle corresponds to $ka \approx 6.9$, and one additional beam is present, at the angle $\theta_r \approx -25^\circ$. The ratio $|I_{-2}/I_{-1}|$ at this value of ka is approximately 0.14, so that the power radiated by the additional beam is only about 2 per cent of that radiated by the desired beam. Since this ratio is proportional, to first order, to the modulation depth M , the ratio of radiated powers would be reduced by a factor of about four if M had

been chosen to be 0.2, rather than 0.4. The attenuation constant α thus represents essentially the leakage into the desired beam only. The radiation pattern may then be determined by standard techniques; if the antenna is long and the leakage is slow, a very narrow beam can be obtained.

The motivation for this study was related to forward end-fire radiation, *i.e.*, $\theta_r = 90^\circ$ in Fig. 13. The same considerations discussed above for $\theta_r = 30^\circ$ apply to the end-fire case, except for the validity of the perturbation expression (18), which affects only the detailed numerical values. Thus, the high-gain end-fire antenna of the modulated surface wave type is in reality a form of *leaky wave* antenna. With a long structure designed for slow leakage, a high end-fire gain is seen to be achievable.

ACKNOWLEDGMENT

The authors wish to thank F. J. Zucker of the Air Force Cambridge Research Center for many stimulating discussions on the subject of modulated surface-wave antennas.

On the Excitation of the Waves of Proper Solutions

KOICHI FURUTSU†

Summary—An exact and explicit solution is obtained for an electro-magnetic field excited by a vertical electric dipole located over a flat plane of arbitrary surface impedance Z . Though there are four asymptotic expansions, the expansion by Hadamard's method seems to have the most explicit physical meaning. According to this expansion only the proper solution of the ordinary surface wave can be present explicitly on the established condition of $\arg(Z) > \pi/4$.

The variation of the field on a spherical surface due to the change of surface impedance is investigated by the use of the ordinary B. Van der Pol and H. Bremmer formula. In this case, when $\arg(Z) > \pi/3$ and $|Z|$ is larger than some definite value $|Z_0|$, one of the terms of the formula is found to have the asymptotic form similar to the surface wave term in a flat surface case and, when the surface is highly inductive, it becomes the leading term at large distance from the dipole. The height-gain factor of this term first decreases rapidly with the height up to some point and then gradually increases. The growing up at large height is due to the radiation field. On the other hand, when $|Z| \ll |Z_0|$, the leading term can scarcely have the correspondence with that in a flat surface case.

The propagation of surface wave over a spherical surface across several boundaries of discontinuity of the surface impedance is considered. The leading term is found to take the simple form as in the ordinary transmission line. The results in the case of flat plane are derived from these results as the asymptotic forms in the limit of infinite radius of curvature of spherical surface.

The physical meaning of the proper solution such as ordinary surface wave is discussed. The physical condition is not the support condition but rather the condition on which, in some domain of space, the proper solution becomes the leading term of wave by an excitation whose dimension is of the order of magnitude of the wavelength or smaller. As in the case of spherical surface, the radiation field and the surface wave field are generally inseparable. The physical wave will be the leading term of the excited wave itself which may sometimes be a proper solution, such as ordinary surface wave, and sometimes be a mixture with the radiation field. The set of proper solutions is not unique.¹ The Green function of the electromagnetic field constitutes the infinite sets of proper solutions of a continuous spectrum, by each set of which the observed field at a given point can be represented, and in each set of which some members of solutions have the properties of surface wave if there is a highly inductive surface.

The field strength excited by a dipole is investigated as a function of the surface impedance in the case where the surface is a flat plane (Section I) or a spherical surface (Section II). The results in the latter case are then applied to the surface wave propagating across several boundaries of discontinuity of surface impedance (Section III). The physical meaning of the waves of proper solutions, such as the surface wave, is finally discussed in Section IV.

† Radio Res. Labs., Kokubunji, Tokyo, Japan.

¹ The definition of "proper solution" used here is somewhat different from that used in the ordinary eigen-value problem of wave equation. Here, a solution is proper if it satisfies the boundary condition of E-M fields on all the given boundaries and, simultaneously, the outward propagating wave conditions at infinity. The proper solutions thus defined must be singular at least at a point in space, otherwise they would vanish identically in the opened system. This situation is the same as in the ordinary surface wave, such as the excited Sommerfeld surface wave on a flat plane or cylindrical wire, which is indeed singular on the line or plane passing through the exciting point.

I. STATEMENT OF THE PROBLEM AND EXACT SOLUTION IN THE CASE OF FLAT SURFACE

WE consider here the field excited by an electric dipole orientated in the direction vertical to a flat surface. Using the cylindrical coordinates $(x) = (r, \varphi, z)$ having the origin on the surface and the z -axis in the outward normal direction passing through the dipole, each component of the field is given by

$$\begin{aligned} E_z &= \frac{1}{ik} \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial}{\partial r} \Psi \right), \\ E_r &= \frac{1}{ik} \sqrt{\frac{\mu_0}{\epsilon_0}} \frac{\partial^2}{\partial z \partial r} \Psi, \\ H_\varphi &= -\frac{\partial}{\partial r} \Psi, \quad k = \sqrt{\epsilon_0 \mu_0} \omega. \end{aligned} \quad (1)$$

Here, Ψ is the solution of the equation

$$(\Delta + k^2)\Psi(x) = -\delta(x - x_1), \quad (2)$$

with the boundary condition

$$Z = -(E_r/H_\varphi)_{z=0}, \quad (3)$$

and the outward propagating wave condition at $z \rightarrow +\infty$. In (2), the point $x_1(0, z_1)$ is the position of the dipole, and the Z in (3) is the ordinarily defined surface impedance.

The exact solution of E_z at the point $x_2(r, \varphi, z_2)$, say $E_z(x_2, x_1)$, takes the form

$$\begin{aligned} E_z(x_2, x_1) &= E_z^0(x_2, x_1) + \frac{k}{4\pi i R} e^{-ikR} \left[\frac{(3 \sin^2 \theta - 1)}{(kR)^2} \right. \\ &\quad + i \frac{3 \sin^2 \theta - 1 - 2 \sin \zeta \sin \theta}{kR} \\ &\quad + \cos^2 \theta \left\{ \frac{\sin \theta - \sin \zeta}{\sin \theta + \sin \zeta} \right\} \\ &\quad \left. + \frac{2 \cos^2 \zeta \sin \zeta}{\sin \theta + \sin \zeta} F \right]. \end{aligned} \quad (4)$$

Here, $E_z^0(x_2, x_1)$ is the field strength in free space and

$$\begin{aligned} Z &= \sqrt{\mu_0/\epsilon_0} \sin \zeta, \quad r = R \cos \theta, \quad z_1 + z_2 = R \sin \theta, \\ R &= \sqrt{r^2 + (z_1 + z_2)^2}. \end{aligned} \quad (5)$$

The factor F in (4) is given by the integral representation very similar to that of Bessel function (Appendix I),

$$F = 1 - ikR(\sin \theta + \sin \zeta) \int_{\psi_0}^{\infty - i\pi/2} \exp [ikR \cos \theta \cos \zeta \{ \cosh \psi_0 - \cosh \psi \}] d\psi, \quad (6)$$

where

$$\cosh \psi_0 = \frac{1 + \sin \theta \sin \zeta}{\cos \theta \cos \zeta}, \quad \sinh \psi_0 = \frac{\sin \theta + \sin \zeta}{\cos \theta \cos \zeta}. \quad (7)$$

The integration path in (6) can be divided into the two parts from ψ_0 to 0 and from 0 to $\infty - i\pi/2$. The contribution from the former part can be expanded in power series of kR . For that of the latter part, we can use the relation

$$H_\nu^{(2)}(z) = -\frac{2}{\pi i} e^{1/2\nu\pi i} \int_0^{\infty - i\pi/2} e^{-iz \cosh \psi} d\psi, \quad |\arg(z)| < \frac{\pi}{2}. \quad (8)$$

Thus,

$$F = 1 + ikR(\sin \theta + \sin \zeta) \cdot \left[i \frac{\pi}{2} \exp [ikR(1 + \sin \theta \sin \zeta)] H_0^{(2)}(kz \cos \theta \cos \zeta) + \sum_{n=0}^{\infty} C_n \{ ikR \cos \theta \cos \zeta \}^n \right]. \quad (9)$$

Here,

$$C_n = \frac{1}{n!} \int_0^{\psi_0} \{ \cosh \psi_0 - \cosh \psi \}^n d\psi, \quad (10)$$

which has the recurrence relation

$$C_n = \frac{1}{n^2} \{ (2n-1) \cosh \psi_0 C_{n-1} - \sinh^2 \psi_0 C_{n-2} \}, \quad (11)$$

with

$$C_0 = \psi_0, \quad C_1 = \psi_0 \cosh \psi_0 - \sinh \psi_0. \quad (12)$$

In the special case of $|\psi_0| \ll 1$,

$$C_n \simeq \frac{1}{1 \cdot 3 \cdot 5 \cdots (2n+1)} \psi_0^{2n+1}. \quad (13)$$

In the case where $\theta \ll 1$, $|\zeta| \ll 1$ and $kR \gg 1$, it follows from (5) and (7) that $\psi_0 \simeq \theta + \sqrt{(\epsilon_0/\mu_0)}Z$ and the function F takes the approximate form

$$F \simeq 1 - ikR\psi_0 \int_{\psi_0}^{\infty} \exp [i\frac{1}{2}kR(\psi_0^2 - \psi^2)] d\psi = 1 - i2\sqrt{\rho}e^{-\rho} \int_{i\sqrt{\rho}}^{\infty} e^{-t^2} dt, \quad \rho = -i\frac{1}{2}kR\psi_0^2, \quad (14)$$

which is the formula ordinarily used.

The asymptotic expansions of the function F are generally divergent series. However, by the use of Hadamard's method, it is possible to modify the asymptotic expansions, so that they become convergent series with a negligible remainder term. Since the problem of the remainder in the asymptotic expansions does not occur, these expansions seem to be most important. We shall first obtain it and then the ordinary expansions by the familiar method.

From (62) in Appendix I, we have by the change of variable $\phi = u + \phi_0$ and by the help of (5) and (7) the other integral representation of F :

$$F = 1 - \int_0^{\infty \exp i\beta} e^{-u} \left\{ 1 - \frac{i u}{2kR \sin^2 \frac{1}{2}(\theta + \zeta)} \right\}^{-1/2} \cdot \left\{ 1 - \frac{i u}{2kR \cos^2 \frac{1}{2}(\theta - \zeta)} \right\}^{-1/2} du, \quad \frac{\pi}{2} > \beta > -\frac{\pi}{2}, \quad |\arg [\sin \frac{1}{2}(\theta + \zeta)]| \leq \frac{\pi}{2}, \quad |\arg [\cos \frac{1}{2}(\theta - \zeta)]| \leq \frac{\pi}{4}. \quad (15)$$

Here, the integration path of u is taken on the upper side of all the branch points, as is indicated in Fig. 6.

$$\left\{ 1 - \frac{i u}{2kR \sin^2 \frac{1}{2}(\theta + \zeta)} \right\}^{-1/2} \left\{ 1 - \frac{i u}{2kR \cos^2 \frac{1}{2}(\theta - \zeta)} \right\}^{-1/2} = \sum_{n=0}^{\infty} \frac{b_n}{n!} \left(\frac{u}{-ikR} \right)^n, \quad (16)$$

whose right side is a uniformly convergent series in the domain of $|u| \leq 2kR |\sin^2 \frac{1}{2}(\theta + \zeta)|$. Hence, when all the branch points are on the lower side of the real axis of u , (15) can be expressed by the convergent series and the negligible remainder term:

$$F = - \sum_{n=1}^{\infty} \frac{b_n \gamma(n+1, 2kR |\sin^2 \frac{1}{2}(\theta + \zeta)|)}{n! (-ikR)^n} + O[\exp(-2kR |\sin^2 \frac{1}{2}(\theta + \zeta)|)], \quad \arg [\sin \frac{1}{2}(\theta + \zeta)] \leq \frac{\pi}{4}. \quad (17)$$

Here, $\gamma(n, x)$ denotes the incomplete Gamma-function of Legendre

$$\gamma(n, x) = \int_0^x u^{n-1} e^{-u} du, \quad (18)$$

and

$$b_n = \frac{n!}{(-2)^n} \sum_{l=0}^n \binom{-\frac{1}{2}}{l} \binom{-\frac{1}{2}}{n-l} \{ \sin \frac{1}{2}(\theta + \zeta) \}^{-2l} \cdot \{ \cos \frac{1}{2}(\theta - \zeta) \}^{-2(n-l)}, \quad (19)$$

or

$$\begin{aligned} b_1 &= 2^{-2} \left\{ \sin^{-2} \frac{1}{2}(\theta + \zeta) + \cos^{-2} \frac{1}{2}(\theta - \zeta) \right\}, \\ b_2 &= 2^{-4} \left\{ 3 \sin^{-4} \frac{1}{2}(\theta + \zeta) \right. \\ &\quad \left. + 2 \sin^{-2} \frac{1}{2}(\theta + \zeta) \cos^{-2} \frac{1}{2}(\theta - \zeta) + 3 \cos^{-4} \frac{1}{2}(\theta - \zeta) \right\}, \\ b_3 &= 3 \cdot 2^{-6} \left\{ 5 \sin^{-6} \frac{1}{2}(\theta + \zeta) + 3 \sin^{-4} \frac{1}{2}(\theta + \zeta) \cos^{-2} \frac{1}{2}(\theta - \zeta) \right. \\ &\quad \left. + 3 \sin^{-2} \frac{1}{2}(\theta + \zeta) \cos^{-4} \frac{1}{2}(\theta - \zeta) \right. \\ &\quad \left. + 5 \cos^{-6} \frac{1}{2}(\theta - \zeta) \right\}, \dots, \end{aligned} \quad (20)$$

$$b_n \simeq 1 \cdot 3 \cdot 5 \cdots (2n-1)(\theta + \zeta)^{-2n}, \quad \theta \ll 1, \quad |\zeta| \ll 1. \quad (21)$$

When the one of branch point is on the upper side of the real axis, we must add to (17) the contribution by the contour C_1 shown in Fig. 6 which can be expressed by the ordinary Hankel function. Thus,

$$\begin{aligned} F &= -\pi kR(\sin \theta + \sin \zeta) \\ &\quad \cdot \exp [ikR(1 + \sin \theta \sin \zeta)] H_0^{(2)}(kR \cos \theta \cos \zeta) \\ &\quad - \sum_{n=1}^{\infty} \frac{b_n \gamma(n+1, 2kR |\sin^2 \frac{1}{2}(\theta + \zeta)|)}{n!(-ikR)^n} \\ &\quad + O[\exp(-2kR |\sin^2 \frac{1}{2}(\theta + \zeta)|)], \\ &\quad \arg [\sin \frac{1}{2}(\theta + \zeta)] \geq \pi/4. \end{aligned} \quad (22)$$

It is impossible for the two branch points to be both on the upper side of the real axis. The inequality in (22) gives the support condition of surface wave and, when $|\zeta| \ll 1$, it gives the same condition as J. R. Wait derived for the asymptotic expansions of (14).²

On the other hand, the ordinary asymptotic expansions are also obtained by the use of (16). The results are as follows:

$$(i) \quad \pi/2 > \arg [\sin \frac{1}{2}(\theta + \zeta)]$$

$$F \sim - \sum_{n=1}^{\infty} b_n (-ikR)^{-n}. \quad (23)$$

$$(ii) \quad \arg [\sin \frac{1}{2}(\theta + \zeta)] > 0, \quad \arg [\sin \frac{1}{2}(\theta + \zeta)] > \arg [\cos \frac{1}{2}(\theta - \zeta)]$$

$$\begin{aligned} F &\sim -\pi kR(\sin \theta + \sin \zeta) \\ &\quad \cdot \exp [ikR(1 + \sin \theta \sin \zeta)] H_0^{(2)}(kR \cos \theta \cos \zeta) \\ &\quad - \sum_{n=1}^{\infty} b_n (-ikR)^{-n}. \end{aligned} \quad (24)$$

$$(iii) \quad \arg [\sin \frac{1}{2}(\theta + \zeta)] > 0, \quad \arg [\cos \frac{1}{2}(\theta - \zeta)] > 0$$

$$\begin{aligned} F &\sim -2\pi kR(\sin \theta + \sin \zeta) \\ &\quad \cdot \exp [ikR(1 + \sin \theta \sin \zeta)] J_0(kR \cos \theta \cos \zeta) \\ &\quad - \sum_{n=1}^{\infty} b_n (-ikR)^{-n}. \end{aligned} \quad (25)$$

$$\begin{aligned} (iv) \quad &\arg [\cos \frac{1}{2}(\theta - \zeta)] > 0, \\ &\arg [\cos \frac{1}{2}(\theta - \zeta)] > \arg [\sin \frac{1}{2}(\theta + \zeta)] \\ &F \sim \pi kR(\sin \theta + \sin \zeta) \\ &\quad \cdot \exp [ikR(1 + \sin \theta \sin \zeta)] H_0^{(1)}(kR \cos \theta \cos \zeta) \\ &\quad - \sum_{n=1}^{\infty} b_n (-ikR)^{-n}. \end{aligned} \quad (26)$$

Here, b_n 's are the same as in (20).

It may be remarked that, when $\theta \neq 0$, there is a wide domain of ζ in which the asymptotic expansions (i)–(iii) are simultaneously valid.

In the following, we shall briefly examine the expansion (22) in the case in which $kR \gg 1$, $\theta \ll 1$ and $|Z| \ll \sqrt{\mu_0/\epsilon_0}$. If $\frac{1}{2}kR|\theta + Z\sqrt{\epsilon_0/\mu_0}|^2 \gg 1$, the first term on the right side of (22) then takes the asymptotic form

$$\begin{aligned} &-\sqrt{2\pi kR}(\theta + Z\sqrt{\epsilon_0/\mu_0}) \\ &\quad \cdot \exp \left[i \left\{ \frac{1}{2} kR(\theta + Z\sqrt{\epsilon_0/\mu_0})^2 + \frac{\pi}{4} \right\} \right], \end{aligned} \quad (27)$$

whose magnitude, when the impedance Z is purely inductive, is $\sqrt{2\pi kR\epsilon_0/\mu_0} Z$ for $\theta=0$ and decreases exponentially with θ . Thus, (27) has the properties of surface waves, as is well known.

II. VARIATION OF THE FIELD OVER A SPHERICAL SURFACE DUE TO THE CHANGE OF SURFACE IMPEDANCE

In this section, it is intended not to derive the exact field strength as in the preceding section, but to see the numerical changes of this strength and the height-gain function due to the change of surface impedance. Therefore, the formula of B. Van der Pol and H. Bremmer is used here in a modified form.³ Thus, when the attenuation coefficient A is defined by the ratio of the field strength to twice the field strength in free space, we have

$$\begin{aligned} A &= \sum_{s=0}^{\infty} (2\pi k r)^{1/2} (ka)^{-1/3} (2\tau_s - 1/\delta^2)^{-1} g_{\tau_s}(z_1) g_{\tau_s}(z_2) \\ &\quad \cdot \exp [-i \{ k r (ka)^{-2/3} \tau_s + \pi/4 \}], \\ \delta &= -iZ^{-1} \sqrt{\frac{\mu_0}{\epsilon_0}} (ka)^{-1/3}. \end{aligned} \quad (28)$$

Here, a is the radius of curvature of the surface, Z the surface impedance over which the waves propagate, r the propagation distance measured along the spherical surface, z_1 and z_2 the heights of the transmitting and receiving points from the surface, and

$$g_{\tau_s}(z) = \frac{(2kz(ka)^{-1/3} - 2\tau_s)^{1/2} H_{1/3}^{(2)} \left[\frac{1}{3} \{ 2kz(ka)^{-1/3} - 2\tau_s \}^{3/2} \right]}{(-2\tau_s)^{1/2} H_{1/3}^{(2)} \left[\frac{1}{3} (-2\tau_s)^{3/2} \right]}, \quad (29)$$

² J. R. Wait, "Excitation of surface waves on conducting, stratified, dielectric-clad, and corrugated surfaces," *J. Res. Natl. Bur. Standards*, vol. 59, pp. 365–377; April, 1957.

³ J. R. Wait, "Radiation from a vertical antenna over a curved stratified ground," *J. Res. Natl. Bur. Standards*, vol. 56, pp. 237–244; August, 1956. (Formulation in terms of surface impedance.)

where τ_s 's are the set of roots of the equation

$$H_{1/3}^{(2)}(x) - \delta(3x)^{1/3} e^{-i2\pi/3} H_{2/3}^{(2)}(x) = 0, \\ x = \frac{1}{3}(-2\tau_s)^{3/2}. \quad (30)$$

In the following, we shall investigate the roots of (30) as the functions of the surface impedance Z . Eq. (30) approaches $H_{2/3}^{(2)}(x) = 0$ as $|\delta| \rightarrow \infty$. Hence, τ_s 's in the general case can be expanded in power series of δ^{-1} as follows:

$$\tau_s = \tau_s' - \frac{1}{2\tau_s'} \delta^{-1} - \frac{1}{8\tau_s'^3} \delta^{-2} - \left(\frac{1}{16\tau_s'^5} + \frac{1}{12\tau_s'^2} \right) \delta^{-3} \\ + \dots, \quad s = 0, 1, 2, \dots, \quad (31)$$

which is convergent (see Appendix II), in the ranges of Z

$$|Z| < |Z_0|, \quad s = 0, \quad |Z| < |Z_{s-1}|, \quad s \geq 1. \quad (32)$$

Here, τ_s 's are the set of roots of $H_{2/3}^{(2)}(x) = 0$; the quantities Z_s are defined by (69) in Appendix II.

On the other hand, as $\delta \rightarrow 0$, (30) generally approaches the two different equations: one is $H_{1/3}^{(2)}(x) = 0$ and the other is

$$\delta(3x)^{1/3} = e^{i2\pi/3} H_{1/3}^{(2)}(x) / H_{2/3}^{(2)}(x). \quad (33)$$

In the limit $x \rightarrow \infty \exp(i\beta)$ with $\pi > \beta > -2\pi$.

Eq. (33) takes the asymptotic form

$$-i\delta(3x)^{1/3} \sim 1 + \frac{i}{6x} - \frac{5}{144} \frac{1}{x^2} - i \frac{5}{648} \frac{1}{x^3} + \dots, \quad (34)$$

which gives the root, say $\tau_s = \tau_0$,

$$\tau_0 \sim \frac{1}{2\delta^2} \left\{ 1 - \delta^3 - \frac{5}{8} \delta^6 - \frac{11}{12} \delta^9 + \dots \right\}, \\ |\delta| \ll 1, \quad (35)$$

$$2\pi - \pi/6 > \arg(\delta^{-1}) > \pi - \pi/6,$$

or

$$4\pi/3 > \arg(Z) > \pi/3.$$

The other roots, say τ_s 's ($s=1, 2, \dots$), can be expanded in power series of δ which are, as $\delta \rightarrow 0$, reduced to the roots τ_s'' 's ($s=0, 1, 2, \dots$) of $H_{1/3}^{(2)}(x) = 0$, respectively. Thus, on reference to Appendix II,

$$\tau_s = \tau_{s-1}'' - \delta - \frac{2}{3}\tau_{s-1}''\delta^3 + \frac{1}{2}\delta^4 + \dots, \\ s = 1, 2, 3, \dots, \quad (36)$$

$$|Z| > |Z_{s-1}|, \quad \arg(Z_{-s+1}) > \arg(Z) > \arg(Z_{s-1}).$$

Here, $\arg(Z_{+s}) \sim \pi/3$ and $\arg(Z_{-s}) \sim 4\pi/3$.

In the case where $\arg(Z)$ is in the range of (35), τ_s 's of $s=1, 2, 3, \dots$, in (31) and (36) are the same analytical functions of δ , respectively, and (35) gives the asymptotic expansion of τ_0 in (31) in power series of δ .

On the other hand, in the case of $\pi/3 > \arg(Z)$, (35) does not hold and (36) is connected with (31) by the

replacement of τ_{s-1}'' on the right side by τ_s'' with $s=0, 1, 2, \dots$:

$$\tau_s = \tau_s'' - \delta - \frac{2}{3}\tau_s''\delta^3 + \frac{1}{2}\delta^4 + \dots, \\ s = 0, 1, 2, \dots,$$

$$\arg(Z_{+s}) > \arg(Z) > \arg(Z_{-s}) - 2\pi. \quad (37)$$

When (35) is valid, we have by the use of (29) and the asymptotic form of Hankel function

$$g_{\tau_0}(z) \sim \exp \left[-\frac{1}{3}\delta^{-2} \left\{ (1 - 2kz(ka)^{-1/3}\delta^2)^{3/2} - 1 \right\} \right] \\ \simeq \exp \left\{ i(kz)Z \sqrt{\frac{\epsilon_0}{\mu_0}} \right\}, \quad kz(ka)^{-1/3} |\delta^2| \ll 1. \quad (38)$$

In the same way, by the use of (35), the corresponding term in A of (28) takes the form

$$-\sqrt{2\pi k r \epsilon_0 / \mu_0} Z e^{i\pi/4} \exp \left[ik(z_1 + z_2)Z \sqrt{\frac{\epsilon_0}{\mu_0}} \right. \\ \left. + \frac{i}{2} k r Z^2 \left(\frac{\epsilon_0}{\mu_0} \right) \left\{ 1 - \delta^3 - \frac{5}{8} \delta^6 - \frac{11}{12} \delta^9 - \dots \right\} \right]. \quad (39)$$

If the power series of δ in (39) are omitted, it is the same as (27), since $\theta = (z_1 + z_2)/R \rightarrow 0$ as $R \rightarrow \infty$. Thus, the contribution of the term (39), i.e., that of the term of τ_0 to the field strength, is in agreement with that of the surface wave term in the case of flat plane.

However, the correspondence is not exact and, in Fig. 1, the imaginary part of τ_0 is displayed as a function of δ . Here, according to the notations by K. A. Norton,

$$K = |\delta| = |Z|^{-1} \sqrt{\frac{\mu_0}{\epsilon_0}} (ka)^{-1/3}, \\ b = \frac{3\pi}{2} + 2 \arg(\delta) = \frac{\pi}{2} - 2 \arg(Z). \quad (40)$$

For the numerical calculation (31) is used for $K \gtrsim 1$ and (35) for $K \ll 1$.

As is seen from this figure, the diffraction loss occurs even if the impedance Z is purely inductive ($b = -90^\circ$) for the impedance in the range of $K \gtrsim 1$.

In Fig. 2, the height gain $|g_{\tau_0}(z)|$ is displayed as a function of $kz(ka)^{-1/3}$ and K for $b = -90^\circ$. The magnitude rapidly decreases with z up to about $kz(ka)^{-1/3} \sim K^{-2}$ and then gradually increases, while that of (27) decreases exponentially with z . The growing up at large height in the former is clearly due to the radiation field. Thus we may conclude that, although the term of τ_0 in the series (28) predominantly has the properties of surface waves in the vicinity of the surface on the condition of $\arg(Z) > \pi/3$ and $K \ll 1$, it also includes a contribution to the radiation field, or it expresses the observed field itself at large distance from the transmitter. On the other hand, the term of τ_0 can scarcely have the correspondence with the surface wave on flat plane on the other condition of Z .

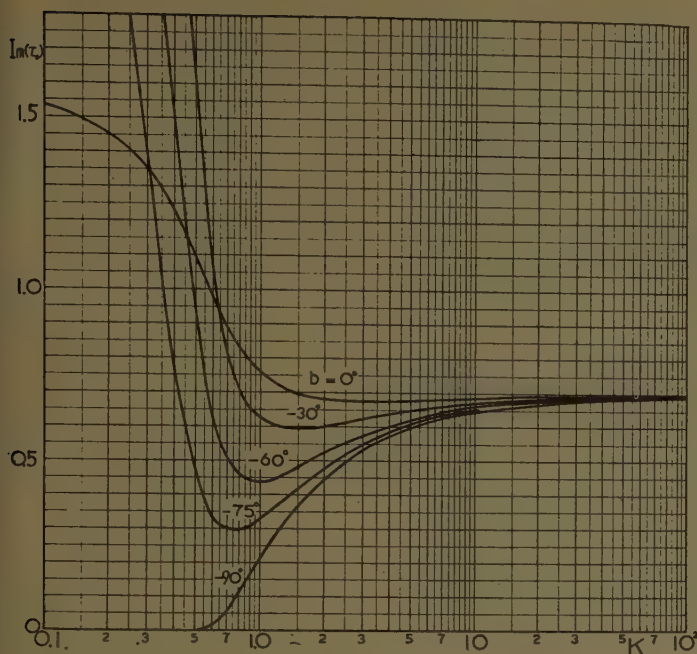


Fig. 1.

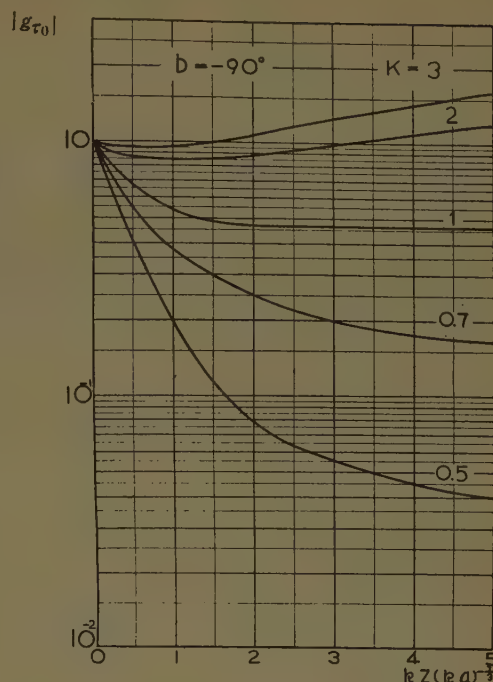


Fig. 2.

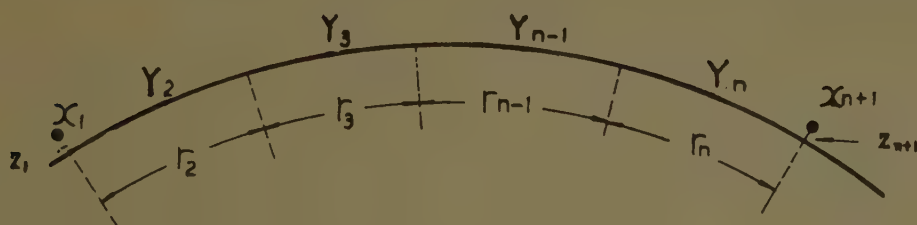


Fig. 3.

III. PROPAGATION OF SURFACE WAVE ACROSS SEVERAL BOUNDARIES OF DISCONTINUITY OF SURFACE IMPEDANCE

By the use of the results in Section II, we shall briefly summarize the results in the case where the waves propagate over several sections of different highly-inductive impedance. A problem of this kind has been investigated in the question of wave propagation over inhomogeneous earth, and several useful formulas were obtained.

As is illustrated in Fig. 3, we consider a general case where the waves propagate over a spherical surface of the radius a . Along the wave path from the point x_1 to x_{n+1} , the surface impedance discontinuously takes the values $Y_2, Y_3, Y_4, \dots, Y_n$ successively. Here, the boundaries between these sections are assumed to be parallel with each other.

As in (28), the field strength $E(x_{n+1}, x_1)$ is here put in the form

$$E(x_{n+1}, x_1) = 2A(z_{n+1} | r_n, r_{n-1}, \dots, r_2 | z_1) E_0(x_{n+1}, x_1), \quad (41)$$

$$r = r_n + r_{n-1} + \dots + r_3 + r_2,$$

where $E_0(x_{n+1}, x_1)$ is the field strength in free space, z_{n+1} and z_1 the vertical coordinates of the points x_{n+1} and x_1 , respectively, and r_m the propagation distance over the m th section.

Provided that every r_m is sufficiently large, so that the reflected waves from the boundaries can be neglected, the factor A is given by

$$A(z_{n+1} | r_n, r_{n-1}, \dots, r_2 | z_1) = \sum_{\tau_n, \tau_{n-1}, \dots, \tau_2} (r/r_n)^{1/2} A(z_{n+1} | r_n)_{\tau_n} \cdot T(r_{n-1})_{\tau_n, \tau_{n-1}} T(r_{n-2})_{\tau_{n-1}, \tau_{n-2}} \dots T(r_2)_{\tau_3, \tau_2} g_{\tau_2}(z_1). \quad (42)^4$$

Here, τ_m denotes the set of values τ_s ($s = 0, 1, 2, \dots$) in Section II for the impedance Y_m and

$$A(z_{n+1} | r_n)_{\tau_n} = (2\pi k r_n)^{1/2} (ka)^{-1/3} (2r_n - 1/\delta_n^2)^{-1} g_{\tau_n}(z_{n+1}) \cdot \exp[-i\{kr_n(ka)^{-2/3}\tau_n + \pi/4\}], \quad (43)$$

⁴ K. Furutsu, "Propagation of electro-magnetic waves over the spherical earth across boundaries separating different earth media," *J. Radio Res. Lab.*, (Japan), vol. 2, pp. 345-398; October, 1955; "Wave propagation over an irregular terrain (I), (II)," vol. 4, pp. 135-153, April, 1957; *ibid.*, vol. 4, pp. 349-393; October, 1957.

$$T(r_n)_{\tau_m, \tau_n} = (\delta_m^{-1} - \delta_n^{-1})(\tau_n - \tau_m)^{-1}(2\tau_n - 1/\delta_n^2)^{-1} \cdot \exp[-ikr_n(ka)^{-2/3}\tau_n],$$

$$\delta_n = -iY_n^{-1}\sqrt{\mu_0/\epsilon_0}(ka)^{-1/3}. \quad (44)$$

When every r_m is sufficiently large, the leading term of (42) will become the first term of series, say that of $\tau_m = \tau_m^0$ ($m=2, 3, 4, \dots, n$) where τ_m^0 denotes the τ_0 in (31) or (35) for $Z = Y_m$. Of course, the latter applies to the δ and Z ranges defined by (35). Furthermore, the result in the case of a flat surface may be obtained by putting $ka \rightarrow \infty$ in this leading term, as (27) is derived from (39).

By the use of (35) and (38), we have from (43) and (44)

$$A(z_{n+1}|r_n)_{\tau_n^0}|_{ka \rightarrow \infty} = -(2\pi kr_n)^{1/2} e^{i\pi/4} Y_n \sqrt{\frac{\epsilon_0}{\mu_0}} \cdot \exp \left[ik \left\{ \frac{1}{2} r_n Y_n^2 \frac{\epsilon_0}{\mu_0} + z_{n+1} Y_n \sqrt{\frac{\epsilon_0}{\mu_0}} \right\} \right],$$

$$T(r_n)_{\tau_m^0, \tau_n^0}|_{ka \rightarrow \infty} = \frac{2Y_n}{Y_m + Y_n} \exp \left[i \frac{1}{2} kr_n Y_n^2 \left(\frac{\epsilon_0}{\mu_0} \right) \right]. \quad (45)$$

Thus, the leading term, say A_0 , in the limit $ka \rightarrow \infty$ becomes

$$A_0|_{ka \rightarrow \infty} = -(2\pi kr)^{1/2} e^{i\pi/4} \sqrt{\frac{\epsilon_0}{\mu_0}} \cdot \frac{2^{n-2} Y_n Y_{n-1} \cdots Y_2}{(Y_n + Y_{n-1})(Y_{n-1} + Y_{n-2}) \cdots (Y_3 + Y_2)} \cdot \exp \left[ik \left\{ \frac{1}{2} \left(\frac{\epsilon_0}{\mu_0} \right) (r_n Y_n^2 + r_{n-1} Y_{n-1}^2 + \cdots + r_2 Y_2^2) + \sqrt{\frac{\epsilon_0}{\mu_0}} (z_{n+1} Y_n + z_1 Y_2) \right\} \right]. \quad (46)$$

Eq. (46) is formally derived if we assume the transmission coefficient of wave amplitude from the m th to the $m+1$ th section to be $2Y_{m+1}/(Y_m + Y_{m+1})$. The energy transmission coefficient then becomes $4Y_{m+1}Y_m/(Y_m + Y_{m+1})^2$, provided Y_m and Y_{m+1} are purely inductive.

It may be interesting to consider the case where some of the sections have flanges, as illustrated in Fig. 4. The terms $T(r_{m-1})_{\tau_m, \tau_n}$'s in (42) are then replaced by $T^{(l)}(r_{m-1})_{\tau_m, \tau_n}$'s defined by

$$T^{(l)}(r_{m-1})_{\tau_m, \tau_{m-1}} = \left\{ \delta_m^{-1} g_{\tau_m}'(a_l - a_m) g_{\tau_{m-1}}(a_l - a_{m-1}) - \delta_{m-1}^{-1} g_{\tau_m}(a_l - a_m) g_{\tau_{m-1}}'(a_l - a_{m-1}) \right\} \cdot \left\{ k(a_{m-1} - a_m)(ka)^{-1/3} + \tau_{m-1} - \tau_m \right\}^{-1} (2\tau_{m-1} - 1/\delta_{m-1}^2)^{-1} \cdot \exp \left[-ikr_{m-1} \left\{ \frac{a_{m-1}}{a} - 1 + (ka)^{-2/3} \tau_{m-1} \right\} \right]. \quad (47)$$

Here, a_l is the vertical coordinate of the top of the flange and

$$\delta_m^{-1} g_{\tau_m}'(z) = k^{-1}(ka)^{1/3} \frac{\partial}{\partial z} g_{\tau_m}(z). \quad (48)$$

By the use of (38), we have

$$g_{\tau_m^0}'(z)|_{ka \rightarrow \infty} = g_{\tau_m^0}(z)|_{ka \rightarrow \infty} = \exp \left\{ i(kz) Y_m \sqrt{\frac{\epsilon_0}{\mu_0}} \right\}. \quad (49)$$

Thus,

$$T^{(l)}(r_{m-1})_{\tau_m^0, \tau_{m-1}^0}|_{ka \rightarrow \infty} = T(r_{m-1})_{\tau_m^0, \tau_{m-1}^0} g_{\tau_m^0}(a_l - a_m) g_{\tau_{m-1}^0}(a_l - a_{m-1})|_{ka \rightarrow \infty}. \quad (50)$$

Hence it follows that the transmission coefficient of wave amplitude from the m th to the $m-1$ th section decreases by the factor

$$\exp [ik \{ (a_l - a_m) Y_m + (a_l - a_{m-1}) Y_{m-1} \} \sqrt{\epsilon_0/\mu_0}]$$

as compared to that in the case without the flange.

IV. THE WAVES OF PROPER SOLUTIONS BY PRACTICAL APPARATUS

The physical waves are those excited by some means; they can be represented, in some cases, by a linear combination of the set of proper solutions and, in other cases, with a linear combination of the set of non-proper solutions. Here, the non-proper solution is preliminarily defined as a solution which satisfies the given boundary conditions in some domain in space but not the conditions in another domain simultaneously. Of course, the waves must not contain a combination of the non-proper solutions which do not satisfy the boundary conditions in the domain including the observing point. According to the way the observing point transfers from one domain to another, the linear combination discontinuously changes from one to another, though the waves are expressed by an analytical function of the observing point in the whole space (except for the exciting point or domain).

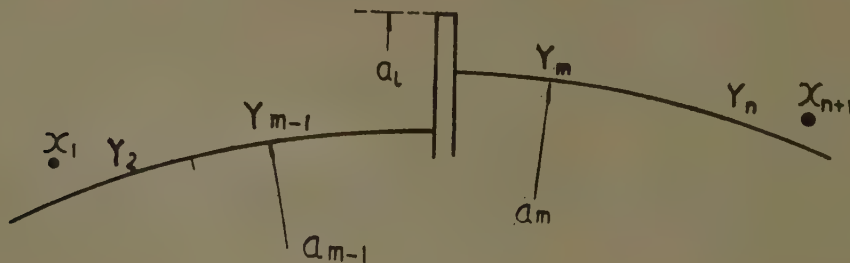


Fig. 4.

The integral representation of the field strength excited by a point source over a flat plane may be a good example of the second case including non-proper solutions. Here, Sommerfeld's surface wave term or the residual term is the proper solution; the remaining term is expressed by the contour integral around the branch cut, and is a linear combination of the non-proper solutions. It may be remarked that the proper term disappears by a suitable deformation of the branch cut; thus, only the non-proper term remains. This fact implies that the proper solution can be represented by the non-proper solutions and the former may not have special physical meaning distinct from the latter (or both are equivalent). This situation will be general in the second case which includes non-proper solutions.

In this way, the physical waves are represented by an infinite series with regard to proper solutions (or an integral in the case of a continuous spectrum) and, in the second case, some series with regard to proper solutions plus an integral with regard to non-proper solutions. Hence, in either case, if the convergence of the series and/or the integrand is sufficiently bad, the properties of the observed waves will be greatly different from those of individual proper solutions.

On the other hand, the restriction on the dimension of excitation, which is probably of the order of the magnitude of wavelength or smaller, will decisively determine the convergence character of the series; the latter character will not be much different from that in the case of a point source excitation. (This point will be discussed later from a more general point of view.) In other words, if some proper solution cannot become the leading term of the wave in any domain of space by the point-source excitation of all kinds, the implication is that this proper solution cannot be an "independent observable," even if it is contained in the expression of field strength explicitly; thus it will not have much physical meaning in itself, though it may have a mathematical meaning.

A good example for this case will be the expression of field strength of (28) over a spherical homogeneous surface, in which the waves are expressed by an infinite series with regard to proper solutions only. In the domain where the conditions $kr(ka)^{-2/3}|\tau_s| \ll 1$ hold for all the proper solutions of $s=0, 1, 2, \dots$, their individual amplitudes decrease with r in proportion to $r^{-1/2}$; thus they have the surface wave-like property. On the other hand, the convergence of series becomes very bad in this domain, and the result of the summation can be sufficiently approximated by the result in the flat surface case. Thus, when both the transmitting and receiving points are on the surface, the actual field has the strength proportional to $F(\rho)r^{-1}$ by (4) and (14), which decreases on the sufficiently lossy surface in proportion to r^{-1} for $|\rho| \ll 1$ and $r^{-1}\rho^{-1} \propto r^{-2}$ for $|\rho| \gg 1$, all of which scarcely express the surface wave-like property.

As is seen in this example, the support condition or the condition upon which some proper solution is con-

tained explicitly in the expression of field strength is not physical condition. The latter will be the condition upon which the proper solution becomes the leading term of field strength in some domain of space. In other words, the physical term is the leading term itself, which may sometimes be a proper solution and sometimes be a mixture with a part of radiation field in ordinary sense.

We have implicitly used the term "proper solution" only for the residual terms of the integral representation of field strength, such as Sommerfeld's surface wave term of arbitrary surface impedance. The latter surface wave indeed satisfies the given boundary condition and the outward propagating wave condition at infinity, but it is singular along the line passing through the exciting point. This singularity is indispensable in the opened system, and thus we are led to generalize the "proper solutions" to include all those solutions satisfying the whole given boundary conditions and the outward propagating wave condition at infinity, irrespective of the singularities in space. In the following, we shall find the infinite sets of "proper solution" of continuous spectrum, by each set of which the field strength at a given observing point can be represented, and from which we can choose the set of most rapid convergence for the integral representation. These sets of "proper solution" are the continuous sets of Green's function of the electro-magnetic field.

In order to see this more explicitly, it is convenient to represent Maxwell's equation in a simple form as follows: since Maxwell's equations are linear differential equations of the first order with regard to E and H , we can represent these equations in Cartesian coordinates in a unified form of

$$\left[\sum_{i=1}^3 \gamma_i \partial_i - k \right] \psi = j, \quad \partial_i = \frac{\partial}{\partial x_i}, \quad (51)$$

with (in Gaussian units)

$$\psi = \begin{pmatrix} iE \\ H \end{pmatrix}, \quad j = \begin{pmatrix} I_e \\ I_m \end{pmatrix}, \quad k = \omega/c. \quad (52)$$

Here, ψ represents the electric and magnetic fields together and thus has the 3×2 components; similarly, j (the electric and magnetic current densities) has these components. γ_i 's are six-rowed matrices and are proved to be Hermitian.

The Green function $\psi(x, x')$ of (51) is here defined as the solution of $j(x) = i\delta(x - x')$, satisfying the whole boundary conditions in space. Here, the coefficient i does not have any important physical meaning. Thus

$$[i(\gamma\partial) - k]\psi(x, x') = i\delta(x - x'),$$

$$(\gamma\partial) = \sum_{i=1}^3 \gamma_i \partial_i. \quad (53)$$

We regard the right side of (53) as if it were to be multiplied by the six-rowed unit matrix. Hence, $\psi(x, x')$ on the left side is also represented by a six-rowed matrix.

$\psi(x', x)$ is proved to satisfy the adjoint equation of

$$\psi(x', x)[-i(\gamma\tilde{\partial}) - k] = i\delta(x - x'). \quad (54)$$

In this case, $\tilde{\partial}_i$'s operate on the coordinates (x) on the left side.

The Green theorem in this formulation becomes as follows: for any continuous vector ψ' and ψ'' , it holds that

$$\begin{aligned} \psi''[i(\gamma\tilde{\partial}) - k]\psi' - \psi''[-i(\gamma\tilde{\partial}) - k]\psi' \\ = i \sum_{i=1}^3 \partial_i(\psi''\gamma_i\psi'), \end{aligned}$$

which gives, on integrating both sides in the space Σ ,

$$\begin{aligned} \int_{\Sigma} \{\psi''[i(\gamma\tilde{\partial}) - k]\psi' - \psi''[-i(\gamma\tilde{\partial}) - k]\psi'\} dv \\ = -i \int_S \psi''(n\gamma)\psi' ds \\ = -i \int [H' \times E'' + E' \times H'']_n ds, \quad (55) \end{aligned}$$

where n is the unit vector inward normal to the surface S enclosing the space Σ , and E' , H' and E'' , H'' are the electric and magnetic fields of ψ' and ψ'' , respectively.

Now, we assume that $\psi(x)$ represents the fields excited at a small domain enclosed by the surface S_1 ; we insert into (55) that: $\psi'(x) = \psi(x)$, $\psi''(x) = \psi(x', x)$ and Σ = the whole space excluding the exciting domain enclosed by S_1 . Then, by the use of (51) and (54), we have the relation

$$\psi(x') = \int_{S_1} \psi(x', x)(n\gamma)\psi(x)ds, \quad (56)$$

because $j(x)$ of $\psi(x)$ vanishes in the space Σ .

Eq. (56) implies that the field ψ at any point x' outside of S_1 can be expressed in terms of the field on S_1 , and also implies that $\psi(x')$ can be represented by the set of "proper solutions" of continuous spectrum $\psi(x', S_1)$, in which S_1 denotes the set of all points on the surface S_1 . On the other hand, the surface S_1 is not unique for a given point x' and thus, if the surface S_2 is another surface enclosing S_1 , and if the point x' is still outside of S_2 , as in Fig. 5, S_1 in (56) can be replaced by S_2 . Therefore,

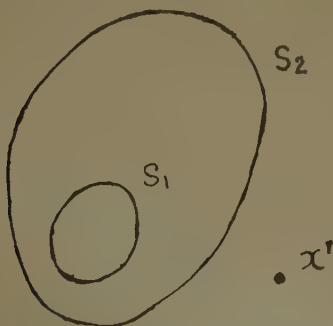


Fig. 5.

it follows that $\psi(x', S_2)$ also constitutes the other set of "proper solutions" by which the field $\psi(x')$ can be represented. Mathematically, the two sets are indeed equivalent⁵ and connected by

$$\psi(x', S_1) = \int_{S_2} \psi(x', x)(n\gamma)\psi(x, S_1)ds, \quad (57)$$

in which $\psi(x', x)$ on the right side belongs to $\psi(x', S_2)$.

Comparing the integral representation of $\psi(x')$ by $\psi(x', S_1)$ of (56) with that by $\psi(x', S_2)$, it is evident that the former will show more rapid convergence than the latter, and that the smaller the surface S_1 becomes, the more approximately each element of $\psi(x', S_1)$ approaches to the observed field $\psi(x')$ itself.

When there is a highly inductive surface, the set of solutions $\psi(x', S_1)$ may be divided into the two groups, one group $\psi(x', S_1')$ having the surface wave properties more than the other, $\psi(x', S_1'')$, where $S_1' + S_1'' = S_1$. It means that $\psi(x', S_1')$ is the set of those waves whose amplitudes first decrease rapidly with the distance of the point x' from the inductive surface in question, and then gradually increase by the ordinary radiation field, as in the case of spherical surface treated in Section III. Therefore, the surface S_1' will be the part of S_1 which is very near to the inductive surface. On the other hand, some elements of $\psi(x', S_1'')$ will scarcely have these properties if their characteristic positions on S_1'' are sufficiently distant from the inductive surface.

The continuous group of solutions $\psi(x', S_1')$ thus obtained are quite equivalent to the ordinarily defined surface waves in that they are the "proper solutions" having the properties of surface wave and belong to $\psi(x', S_1)$ by which the observed field $\psi(x')$ can be represented. The smaller the exciting domain becomes, the smaller the surface S_1 can be taken to be; thus it may happen that S_1 belongs to S_1' or $S_1 = S_1'$, if the exciting domain is chosen at the position sufficiently near to the highly inductive surface in question.

APPENDIX I

The solution of (2) with the boundary condition (3) is given by the well known integral representation which is reduced to the form

$$\begin{aligned} \Psi(x_2, x_1) = & -\frac{i}{4\pi} \int_0^\infty d\lambda \frac{\lambda}{h(\lambda)} J_0(\lambda r) \{ e^{-i h(\lambda) |z_2 - z_1|} \\ & + e^{-i h(\lambda) (z_1 + z_2)} \} \\ & + \frac{i}{2\pi} \int_0^\infty d\lambda \frac{\lambda}{h(\lambda)} J_0(\lambda r) \left\{ \frac{kZ\sqrt{\epsilon_0/\mu_0}}{h(\lambda) + kZ\sqrt{\epsilon_0/\mu_0}} \right\} e^{-i h(\lambda) (z_1 + z_2)}, \\ & h(\lambda) = \sqrt{k^2 - \lambda^2}. \quad (58) \end{aligned}$$

⁵ This statement is true in a restricted sense, because $\psi(x', S_2)$ cannot be represented in linear combination of $\psi(x', S_1)$.

Here, $J_0(z)$ is the Bessel function of the 0th order. For the first two terms on the right side of (58), we can use the formula

$$\frac{1}{R} e^{-ikR} = -i \int_0^\infty d\lambda \frac{\lambda}{h(\lambda)} J_0(\lambda r) e^{-i h(\lambda) z},$$

$$R = \sqrt{r^2 + z^2}. \quad (59)$$

Assuming $\text{Re}(Z) > 0$, it holds, on the whole integration path, that

$$\frac{1}{h(\lambda) + kZ\sqrt{\epsilon_0/\mu_0}}$$

$$= \int_0^\infty \exp[-\{h(\lambda) + kZ\sqrt{\epsilon_0/\mu_0}\}t] dt, \quad (60)$$

by the use of which the last integral in (58) takes the form

$$\frac{i}{2\pi} kZ \sqrt{\frac{\epsilon_0}{\mu_0}} \int_0^\infty dt \exp[-kZ\sqrt{\epsilon_0/\mu_0}t] \int_0^\infty d\lambda \frac{\lambda}{h(\lambda)} J_0(\lambda r)$$

$$\cdot \exp[-ih(\lambda)\{(z_1 + z_2) - it\}],$$

which gives, by the use of (59),

$$-\frac{kZ}{2\pi} \sqrt{\frac{\epsilon_0}{\mu_0}} \int_0^\infty \frac{1}{R} \exp[-k\{iR + Z\sqrt{\epsilon_0/\mu_0}t\}] dt,$$

$$R = \sqrt{r^2 + \{(z_1 + z_2) - it\}^2}. \quad (61)$$

By the change of variable of $\phi = k\{iR + Z\sqrt{\epsilon_0/\mu_0}t\}$, the integral (61) is expressed in a simple form of

$$-i \frac{kZ}{2\pi} \sqrt{\frac{\epsilon_0}{\mu_0}} \int_{\phi_0}^\infty \frac{\exp i\beta}{\sqrt{(\phi - \phi_1)(\phi - \phi_2)}} d\phi. \quad (62)$$

Here, we take $\text{Im}(\phi_1) > \text{Im}(\phi_2)$ and $\sqrt{(\phi - \phi_1)(\phi - \phi_2)} \sim \phi$ as $\phi \rightarrow \infty \exp(i\beta)$ and

$$\phi_1 = ik\{\sqrt{1 - Z^2(\epsilon_0/\mu_0)}r - Z\sqrt{\epsilon_0/\mu_0}(z_1 + z_2)\},$$

$$\phi_2 = ik\{-\sqrt{1 - Z^2(\epsilon_0/\mu_0)}r - Z\sqrt{\epsilon_0/\mu_0}(z_1 + z_2)\},$$

$$\phi_0 = ik\sqrt{r^2 + (z_1 + z_2)^2}, \quad \beta = \arg(1 + Z\sqrt{\epsilon_0/\mu_0}),$$

$$|\arg(Z)| < \pi/2, \quad |\arg\sqrt{1 - Z^2(\epsilon_0/\mu_0)}| < \pi/2; \quad (63)$$

the integration path is taken on the upper side of all the branch points ϕ_1 and ϕ_2 (Fig. 6).

The integral representation (6) in Section I is finally obtained by the change of variable

$$\phi = \frac{1}{2}(\phi_1 + \phi_2) + \frac{1}{2}(\phi_1 - \phi_2) \cosh \psi. \quad (64)$$

APPENDIX II

Eq. (30) in Section II is reduced to

$$f(\beta, \delta) = \chi(\beta) - 2^{1/3} \delta \chi'(\beta) = 0, \quad \chi'(\beta) = \frac{\partial}{\partial \beta} \chi(\beta). \quad (65)$$

Here, $\beta^{3/2} = 3x/2 = \frac{1}{2}(-2\tau_s)^{3/2}$ and

$$\chi(\beta) = \beta^{1/2} H_{1/3}^{(2)}(\frac{2}{3}\beta^{3/2}), \quad \chi''(\beta) + \beta\chi(\beta) = 0. \quad (66)$$

If β_s 's and δ_s 's are the roots of the equations

$$f(\beta, \delta) = 0, \quad f_\beta(\beta, \delta) = \frac{\partial}{\partial \beta} f(\beta, \delta) = 0,$$

then the expansions of $f(\beta, \delta)$ in power series of $(\beta - \beta_s)$ and $(\delta - \delta_s)$ takes the form

$$f(\beta, \delta) = f_s(\beta_s, \delta_s)(\delta - \delta_s) + \frac{1}{2}f_{\beta\beta}(\beta_s, \delta_s)(\beta - \beta_s)^2 + \dots$$

Hence, if $f_{\beta\beta}(\beta_s, \delta_s) \neq 0$, it follows that β is a double value function of δ in the vicinity of $\delta = \delta_s$. Generally speaking, τ_s 's are the values of the multi-value function on the different Riemann planes of δ or Z . These circumstances are shown in Fig. 7.

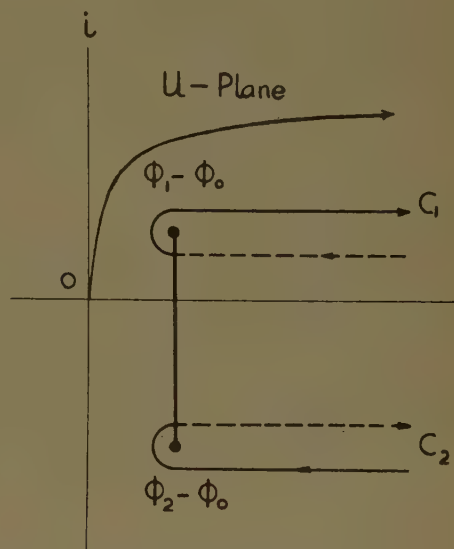


Fig. 6—Integration path of $u = \phi - \phi_0$.

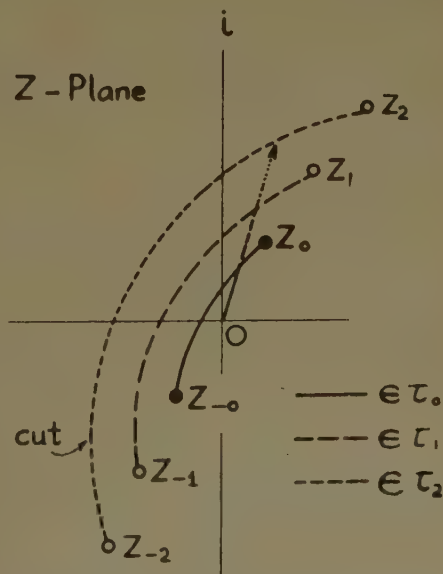


Fig. 7—Positions of branch points on the Riemann planes of Z . The arrow starting from the origin indicates how Z gets into the different Riemann planes with $|Z|$, successively, when $\arg(Z) > \pi/3$.

Using (66),

$$f_{\beta}(\beta, \delta) = (1 + 2^{2/3}\delta^2\beta)\chi'(\beta),$$

which gives $1 + 2^{2/3}\delta_s^2\beta_s = 0$ or

$$\mp i\delta_{\pm s}(3x_{\pm s})^{1/3} = 1, \quad x_s = (x)_{\beta=\beta_s}. \quad (67)$$

Here, according to the sign of the left side, the subscript s^1 is replaced by $\pm s$, respectively. By the substitution of (67) in (30) in Section II, we have

$$H_{1/3}^{(2)}(x_{\pm s}) \mp e^{-i\pi/6}H_{2/3}^{(2)}(x_{\pm s}) = 0, \\ \pi + \alpha \geq \arg (x_{\pm s})^{1/3} \geq \alpha. \quad (68)$$

It is proved that the asymptotic forms of $\tau_{\pm s}$ and $\delta_{\pm s}$, or the corresponding impedance $Z_{\pm s}$ for $s \gg 1$ becomes

$$\tau_{\pm s} \simeq \frac{1}{2}\{3(s+1)\pi\}^{2/3} \left\{ 1 \mp i \frac{\log \{12(s+1)\pi\}}{3(s+1)\pi} \right\} e^{-i\pi/3}, \\ s \gg 1,$$

$$Z_{\pm s} \simeq \pm \sqrt{\frac{\mu_0}{\epsilon_0}} (ka)^{-1/3} \{3(s+1)\pi\}^{1/3} \\ \cdot \left\{ 1 \mp i \frac{\log \{12(s+1)\pi\}}{6(s+1)\pi} \right\} e^{i\pi/3}. \quad (69)$$

Here, s takes all positive integral values.

Surface-Wave Research in Sheffield

M. F. BRACEY†, A. L. CULLEN†, E. F. F. GILLESPIE† AND J. A. STANFORTH†

Summary—The paper is divided into four sections, each dealing with a specific aspect of the current research program in the Department of Electrical Engineering at Sheffield University.

Section I deals with the radiation from a cylindrical surface-wave resonator.

Section II presents some numerical results obtained by an electronic computer of the guide-wavelength/dielectric constant relationship for the EH₁₁ mode on a dielectric rod.

Section III deals with the coupling between two adjacent surface waveguides. Parallel dielectric sheets and parallel rods are both discussed theoretically.

Section IV presents the theory of a new form of surface-wave launching antenna, including an evaluation of its launching efficiency.

I. A CIRCULATING SURFACE-WAVE RESONATOR

Introduction

IN a classical paper on supergain aerials,¹ Chu has calculated the Q -factors associated with weakly-radiating spherical waves. There are, of course, many different ways of calculating and defining Q -factor, but the problem is to choose the most convenient. As Chu points out in his paper, the definition based on energy stored and power dissipated cannot be applied straightforwardly, since the energy stored, and the energy associated with the outward traveling wave cannot easily be separated, at any rate rigorously. Chu overcomes this difficulty by evaluating the Q of an equivalent electrical circuit which he derives from the field equations.

The purpose of this note is to present an alternative approach based on the relationship between Q -factor and the decay time of an oscillatory transient.

The Basic Equation

We shall illustrate the method by applying it in the case of a cylindrical geometry, a case which is of interest in our experimental program.

Referring to Fig. 1, we consider fields which do not depend on z , and which have $E_r = E_\theta = 0$. Such fields can be described outside S by sums of elementary fields of the type

$$E_z = H_n^{(2)}(kr) \cos n\theta \quad (1)$$

$$H_\theta = -j\sqrt{\frac{\epsilon_0}{\mu_0}} H_n^{(2)'}(kr) \cos n\theta \quad (2)$$

where $k = \omega\sqrt{\mu_0\epsilon_0}$.

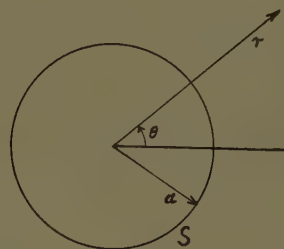


Fig. 1—Illustrating the notation for the cylindrical surface-wave resonator.

The wave impedance looking radially outward at $r = a$ is

$$Z_1 = \left(-\frac{E_z}{H_\theta} \right)_{r=a} = -j\sqrt{\frac{\mu_0}{\epsilon_0}} \frac{H_n(ka)}{H_n'(ka)} \quad (3)$$

in which the superfix (2) of the Hankel functions has been dropped.

Now suppose that the surface S has a capacitive surface reactance, looking *inwards* at $r = a$. We can imagine this to be achieved by a cylinder of infinite permeability coated by a thin shell of material of high dielectric constant. The surface reactance would, in this case, vary with frequency, or k , in the same way as the reactance of a normal capacitor, and so it can be written

$$Z_2 = -j\frac{\omega_c}{\omega}\sqrt{\frac{\mu_0}{\epsilon_0}} = -j\frac{k_c}{k}\sqrt{\frac{\mu_0}{\epsilon_0}} \quad (4)$$

Here ω_c (or k_c) is a real number defining the surface reactance magnitude at a given value of ω (or k).

Both (3) and (4) are valid for complex values of ω corresponding to decaying oscillations, and we can apply the condition for natural oscillations in the simple form given by Schelkunoff,² namely

$$Z_1 + Z_2 = 0. \quad (5)$$

Using (3) and (4) in (5) gives

$$k \frac{H_n(ka)}{H_n'(ka)} = -k_c. \quad (6)$$

This is the basic equation which is to be solved for k .

Approximation Method of Solution

Since k is complex, a direct approach to the solution of (6) is difficult. However, if an approximate value for

† The University of Sheffield, Sheffield, England.

¹ L. J. Chou, "Physical limitations of omni-directional antennas," *J. Appl. Phys.*, vol. 19, pp. 1163-1175; December, 1948.

² S. A. Schelkunoff, "Electromagnetic Waves," D. Van Nostrand Co., Inc., Princeton, N. J., p. 229; 1943.

k can be found for a given n and given k_0 , this value can be corrected to give a more accurate value by using a Taylor series expansion. Thus, if k_1 is a first approximation to k , we write

$$k = k_1(1 + \eta) \quad (7)$$

where η is a small quantity to be determined. Eq. (6) can then be written

$$(1 + \eta)[F(k_1 a) + \eta k_1 a F'(k_1 a)] = -\frac{k_0}{k_1} \quad (8)$$

where

$$F(x) = \frac{H_n(x)}{H_n'(x)}. \quad (9)$$

Neglecting η^2 in (8), we find

$$\eta = -\frac{\frac{k_0}{k_1} + F(k_1 a)}{F(k_1 a) + k_1 a F'(k_1 a)}. \quad (10)$$

The approximate value k_1 can be found by arguing on physical grounds that if the surface were flat with perpendicular end plates, rather than cylindrical, the resonance condition would not be greatly altered. When this value of k_1 is introduced, and when the Hankel functions are introduced, recurrence formulas being used to eliminate derivatives, we find the following explicit formula for η ,

$$\eta = -\frac{\frac{1}{\sqrt{\left(\frac{n^2}{x^2} - 1\right)}} + \frac{2H_n(x)}{H_{n-1}(x) - H_{n+1}(x)}}{x + \frac{4H_n(x)}{H_{n-1}(x) - H_{n+1}(x)} + \left(x - \frac{n^2}{x}\right)\left\{\frac{2H_n(x)}{H_{n-1}(x) - H_{n+1}(x)}\right\}^2}. \quad (11)$$

In this formula, x is the number of free-space wavelengths around the circumference of the cylinder, and n is the number of actual (surface-wave) wavelengths around the cylinder.

An alternative method of arriving at a first approximation depends on the fact that the Y_n and Y_n' functions are considerably greater than the J_n and J_n' functions in (6), so that we can get an approximate solution by writing

$$k_1 \frac{Y_n(k_1 a)}{Y_n'(k_1 a)} \cong -k_0 \quad (6a)$$

and regarding this equation as specifying k_0 for a given k_1 .

When an approximate solution has been found, it can presumably be improved still further by choosing k_0 in order that the real part of η vanishes, so that the frequency of oscillation will be exact.

Determination of Q from η

Write $\omega = \omega' + j\omega''$. The time factor $e^{j\omega t}$ assumed in the foregoing equations then becomes $e^{j\omega' t} e^{-\omega'' t}$ corresponding to a decaying oscillation if ω'' is positive. For a simple RLC circuit,

$$\omega'' = \frac{R}{2L}, \quad \text{and} \quad Q = \frac{\omega' L}{R}.$$

Hence

$$Q = \frac{\omega'}{2\omega''}. \quad (12)$$

To the order of approximation used earlier, we find

$$Q \cong \frac{1}{2\eta''} \quad (13)$$

$$\frac{\delta f}{f_0} \cong \eta'. \quad (14)$$

The last equation gives the shift in frequency due to rolling up a flat surface-wave resonator of length $2\pi a$ into a cylinder of radius a .

Numerical Example

Consider a cylinder of three TEM wavelengths circumference, loaded so that six surface-wave wavelengths are accommodated around its circumference. In our notation $x = 3$, $n = 6$. Substituting in (11), we get

$$\eta' = -0.0202$$

$$\eta'' = +0.00125.$$

Hence from (13) and (14)

$$Q = 392$$

$$\frac{\delta f}{f_0} = -2.02 \text{ per cent.}$$

If, on the other hand, we use (6a) to get an approximate solution, we find $k_0/k_1 = 0.605$, and with this value

$$Q = 298$$

$$\frac{\delta f}{f_0} = -0.20 \text{ per cent.}$$

Choosing $k_c/k_1 = -0.608$ makes η' vanish, and this gives

$$Q = 291$$

$$\frac{\delta f}{f_0} = 0.00.$$

The last solution is presumably the best estimate of Q -factor which can be obtained in this way, and the discrepancy between this value and the original value of 392 indicates that the best possible value of k_c/k_1 should be used in (10).

Conclusion

A method of calculating Q -factors for high-order weakly-radiating modes is presented, and is shown to lead relatively simply to useful approximate formulas for Q -factor.

Resonances of this kind have been found by Wait and Conda³ in studies of the radiation patterns of slots in dielectric coated cylinders, and have been shown to be important in modifying the shadow-zone pattern very markedly.

II. A NUMERICAL STUDY OF THE EH_{11} DIPOLE MODE ON A DIELECTRIC ROD

Introduction

The propagation of electromagnetic waves along a dielectric rod has been studied by many authors since the original paper by Hondros and Debye.⁴

A very satisfactory account of the complete theory is given by Stratton⁵ and leads to the following transcendental equation for the particular case of the EH_{11} mode

$$r^2 \left(\frac{1}{u^2} - \frac{1}{v^2} \right)^2 = \left\{ \frac{J_1(u)}{uJ_1(u)} - \frac{H_1^{(1)'}(v)}{vH_1^{(1)'}(v)} \right\} \cdot \left\{ \frac{\kappa J_1'(u)}{uJ_1(u)} - \frac{H_1^{(1)'}(v)}{vH_1^{(1)'}(v)} \right\} \quad (15)$$

where

$$u = \frac{2\pi a}{\lambda_0} (\kappa - r^2)^{1/2}, \quad v = i \frac{2\pi a}{\lambda_0} (r^2 - 1)^{1/2}, \quad r = \frac{\lambda_0}{\lambda_g}. \quad (16)$$

The radius of the rod is a , λ_0 and λ_g are the free-space TEM wavelength and the EH_{11} -mode wavelength, respectively, and κ is the dielectric constant.

³ J. R. Wait and A. M. Conda, "Radiation from slots on dielectric-clad and corrugated cylinders," *J. Natl. Bur. Standards*, vol. 59, pp. 307-316; November, 1957.

⁴ D. Hondros and P. Debye, "Elektromagnetische Wellen an dielektrischen Drahten," *Ann. der Physik*, vol. 32, pp. 465-476; June, 1910.

⁵ J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 524-527; 1941.

Application

Our interest in this mode arose from a proposal to use it as the basis of a technique for the measurement of dielectric constants. The method is to determine λ_g experimentally; then, knowing a and λ_0 , the required dielectric constant κ can be calculated from (15).

Numerical Results

The calculations involved are formidable, as a preliminary study showed, and it became apparent that only by using an electronic computer could the required data be obtained in a reasonable time. Accordingly a program was written by one of the present authors (Gillespie) for a Pegasus computer, and the necessary calculations were performed electronically. We have plotted these data in the form of curves, the ratio λ_g/λ_0 being plotted as ordinate with dielectric constant κ as abscissa, and $2a/\lambda_0$ as a parameter. These curves are shown in Figs. 2 and 3 for a fairly wide range of κ values.

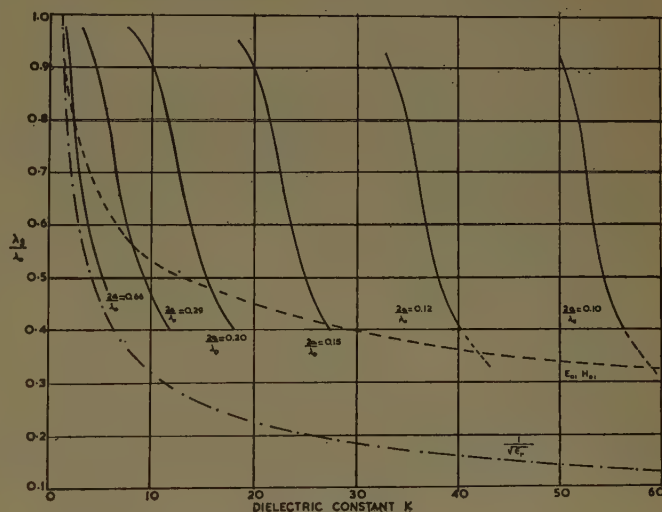


Fig. 2.

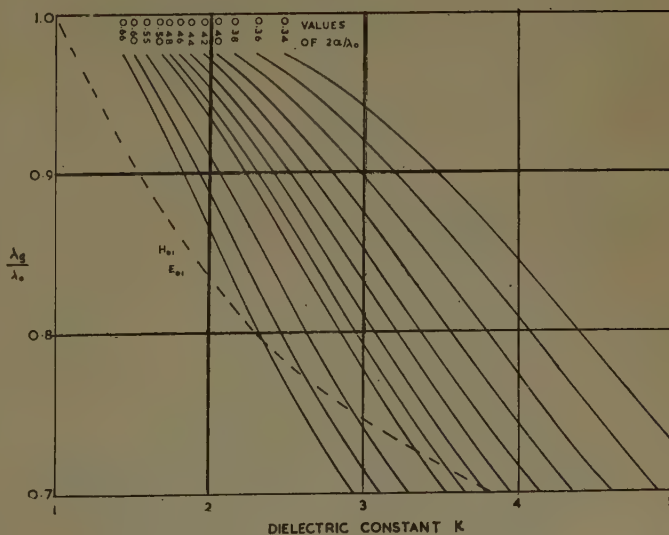


Fig. 3.

It is physically obvious that as $a \rightarrow 0$, $\lambda_g \rightarrow \lambda_0$, and also that as $a \rightarrow \infty$, $\lambda_g \rightarrow \lambda_0/\sqrt{\kappa}$, and these limiting curves are also plotted in Fig. 2.

From a practical point of view, it is more important to know the region of the graph in which the EH_{11} mode is the only possible mode. The next modes to appear as dielectric constant or rod radius is increased, are the E_{01} and H_{01} modes, and the critical condition for both modes corresponds to the first zero of

$$J_0 \left\{ \frac{2\pi a}{\lambda_0} (\kappa - 1)^{1/2} \right\} = 0$$

i.e.,

$$\frac{2\pi a}{\lambda_0} (\kappa - 1)^{1/2} = 2.4048. \quad (17)$$

This condition is also plotted in Fig. 2. Above this curve, only the EH_{11} mode propagates; below it, other modes can propagate.

The calculations presented here represent an extension of previously available results of other workers, to which they are complementary.⁶ In our case, however, the variable has been taken as κ , while previous workers have taken a , the radius of the rod. The reason for our choice of κ is simply one of convenience in the use we wish to make of the results, enabling κ , as the unknown quantity, to be read off at once from the ratio of the measured λ_g and λ_0 values for a given size of rod.

III. COUPLED SURFACE-WAVE GUIDES

Introduction

It is well-known that two transmission lines, continuously coupled along their lengths, exhibit interesting phenomena which have applications in the theory and design of directional couplers.⁷ Usually, the form of coupling convenient in practice is not amenable to analysis in terms of electromagnetic theory.

We have studied two cases in which such analysis is possible. The first case is that of parallel dielectric sheets, and the second case parallel dielectric rods. An exact solution is, of course, possible in the first case, but we have only obtained an approximate solution in the second case.

Parallel Dielectric Sheets

The physical arrangement, notation, and mode configurations are shown in Fig. 4.

As in the case of coupled transmission lines, two normal modes are possible; these are shown in Fig. 4. Each

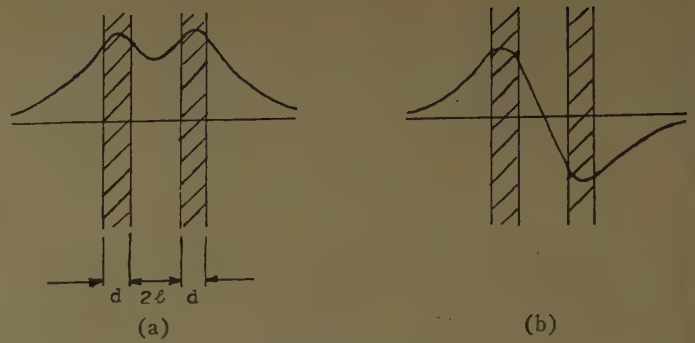


Fig. 4—Mode configuration and notation. (a) symmetric; (b) anti-symmetric.

mode can occur in either TE or TM configuration. The terms symmetric and anti-symmetric as used here refer to E for TE waves and H for TM waves. The phase coefficient β for the various modes can be derived from the following equations:

TE-Waves

$$\frac{q^2 - pq \cot(kdq)}{p^2 + pq \cot(kdq)} = \begin{cases} \tanh(klp); & \text{(symmetric)} \\ \coth(klp); & \text{(anti-symmetric).} \end{cases} \quad (18)$$

TM-Waves

$$\frac{q^2 - \kappa pq \cot(kdq)}{(\kappa p)^2 + \kappa pq \cot(kdq)} = \begin{cases} \tanh(klp); & \text{(symmetric)} \\ \coth(klp); & \text{(anti-symmetric)} \end{cases} \quad (20)$$

where

$$p = \sqrt{\frac{\beta^2}{k^2} - 1} \quad q = \sqrt{\kappa - \frac{\beta^2}{k^2}} \quad p^2 + q^2 = \kappa - 1. \quad (22)$$

These equations have been discussed by Tischer in noting the possible use of modes of this type as low-loss waveguides.⁸

When we consider the case of parallel dielectric rods, an approximate method will be used. This approximate method will first be applied to the parallel sheet case and compared with the exact solution to test its accuracy.

Approximate Calculation Based on the Adiabatic Invariance of Action

In a lossless electromagnetic resonator, action is invariant against a slow deformation.⁹ This can be expressed as follows:

$$\frac{\delta f}{f} = \frac{\delta W}{W}. \quad (23)$$

⁶ W. M. Elsasser, "Attenuation in a dielectric circular rod," *J. Appl. Phys.*, vol. 20, p. 1193; December, 1949. J. W. Duncan and R. H. Du Hamel, "A technique for controlling the radiation from dielectric rod waveguides," Univ. of Ill., Urbana, Ill., Antenna Lab. Rept. No. 11; July 15, 1956.

⁷ S. E. Miller, "Coupled wave theory and waveguide applications," *Bell Sys. Tech. J.*, vol. 33, pp. 661-719; May, 1954.

⁸ F. J. Tischer, "The H-guide, a waveguide for microwaves," 1956 IRE CONVENTION RECORD, pt. 5, pp. 44-47.

⁹ W. R. Maclean, "The resonator action theorem," *Quart. J. Appl. Math.*, vol. 2, p. 329; 1945.

Here, W is the initial energy in the resonator and f is the initial frequency. If the resonator is perturbed slowly so that the energy is increased by δW , the frequency is increased by δf , as given by (23). In particular, if the perturbation leaves the frequency unchanged, there can be no energy change, so

$$\delta f = \delta W = 0. \quad (24)$$

This special case is discussed elsewhere and a simple proof given.¹⁰ Now consider how this can be applied to our problem. Suppose we have a single dielectric sheet supporting a TE wave, and suppose that perfectly-conducting planes (end-plates) are introduced perpendicular to the direction of propagation. There will then be a standing wave in this direction with an integral number of half guide-wavelengths in the distance L between the two end-plates, and a surface wave resonator is formed.

Suppose now that the system is perturbed by bringing in from infinity a perfectly-conducting plane held parallel to the dielectric surfaces, as shown in Fig. 5.

The force F acting on the plane can be calculated from the net radiation pressure, which is

$$p = \frac{1}{2}\mu_0 H_t^2 - \frac{1}{2}\epsilon_0 E_n^2. \quad (25)$$

In a first approximation, we set H_t and E_n equal to twice the corresponding values for the unperturbed field. Integrating F from ∞ to l gives the work δW done on the resonator during this deformation.

In the same way, we can calculate the work done in moving an end-plate through a small distance δL . It is clearly possible to arrange that this end-plate movement exactly cancels the effect of the first deformation considered, so that $\delta W = \delta f = 0$. Clearly $\delta L/L = \delta \lambda_g/\lambda_g$, and so the perturbation of guide wavelength at constant frequency due to putting an electric wall at a distance from the surface waveguide can be found. A comparison of Fig. 5(b) with Fig. 4(b) shows that the perturbed mode we have considered is essentially the same as the anti-symmetric TE mode. For the symmetric TE mode, we bring in a magnetic wall, and the appropriate pressure formula is now

$$p = \frac{1}{2}\epsilon_0 E_t^2 - \frac{1}{2}\mu_0 H_n^2. \quad (26)$$

The same method can be applied to TM modes. Let $\beta = \beta_0 + \delta\beta$, where β_0 is the value of β in the unperturbed case. Thus we find:

TE Wave

$$\frac{\delta\beta}{\beta_0} = \pm \frac{2p^2 q^2 e^{-2klp}}{(\kappa - 1)(1 + p^2)\{2 + kdp\}} \quad (27)$$

TM Wave

$$\frac{\delta\beta}{\beta_0} = \pm \frac{2p^2 q^2 e^{-2klp}}{(\kappa - 1)(1 + p^2)\left\{2 + \frac{kdp}{\kappa} [1 + p^2(\kappa + 1)]\right\}} \quad (28)$$

¹⁰ A. L. Cullen, "A General Method for the Absolute Measurement of Microwave Power," IEE Monograph No. 24; February, 1952.

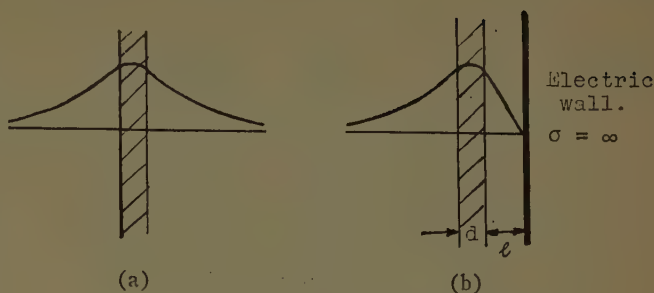


Fig. 5—Perturbation of surface-wave resonator.

Eq. (27) can also be derived from (18) and (19) directly merely by treating e^{-2klp} as a small quantity in these formulas.

Fig. 6 shows the exact and approximate formulas compared. It will be noticed that the agreement is remarkably good, and also that the difference $\lambda_s - \lambda_a$ is more accurate than might be expected from the errors of λ_s and λ_a separately, because the individual errors have opposite signs over most of the range of the values considered.

Dielectric Rods

For rods, the integrations involved in applying the action theorem were carried out approximately using the asymptotic formulas for the Bessel functions.

The unperturbed mode is the EH₁₁ mode so that the previous distinction between TE and TM modes no longer applies.

Fig. 7 shows the configurations of field in the case now considered.

The formulas are as follows:

Vertical Polarization

$$\frac{\delta\beta}{\beta_0} = \pm P \frac{e^{-(2/a)q(l+a)}}{\sqrt{\frac{q}{a}(l+a)}} [XY]. \quad (29)$$

Horizontal Polarization

$$\frac{\delta\beta}{\beta_0} = \pm P \frac{e^{-(2/a)q(l+a)}}{\sqrt{\frac{q}{a}(l+a)}} \quad (30)$$

+ for symmetric, - for antisymmetric mode.

The value β_0 for a single rod is obtained by solving a transcendental equation, as mentioned in Section II. The functions P , X , and Y are given by

$$\begin{aligned} \frac{1}{P} = & \frac{\pi\sqrt{\pi}}{2} (akq)^2 \{H_1^{(1)}(iq)\}^2 \left[\left(\frac{Y+1}{q^2} \right)^2 - \left(\frac{Y+\kappa}{p^2} \right)^2 \right] \\ & + \frac{(\kappa F - G)}{a^2 k^2} + \frac{\beta^2}{2k^2} \left\{ (\kappa + XY) \left(\frac{1}{p^2} + F^2 \right) \right. \\ & \left. + (1 + XY) \left(\frac{1}{q^2} - G^2 \right) \right\} \end{aligned} \quad (31)$$

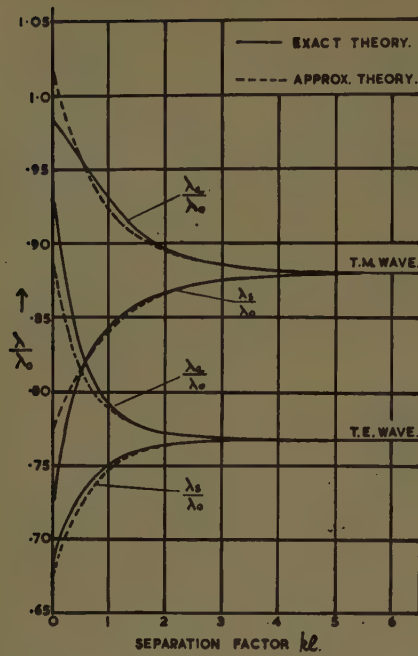


Fig. 6—Variation of λ_s/λ_0 and λ_a/λ_0 with separation factor kl . Sheet thickness factor $kd=\pi/2$. $\kappa=2.56$

Vertical polarization

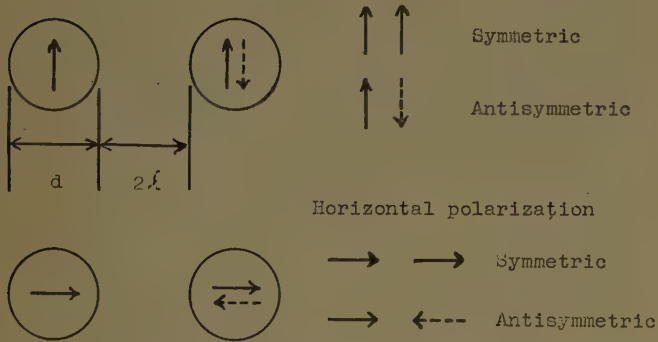


Fig. 7—Notation and nomenclature for rods.

$$\begin{aligned}
 X &= \frac{1}{p^2} + \frac{1}{q^2}, & Y &= \frac{\kappa}{p^2} + \frac{1}{q^2}, \\
 F &= \frac{1}{p} \frac{J_0(p)}{J_1(p)}, & G &= \frac{1}{iq} \frac{H_0^{(1)}(iq)}{H_1^{(1)}(iq)}, \\
 f &= F - \frac{1}{p^2}, & g &= G + \frac{1}{q^2}
 \end{aligned} \quad (32)$$

where p and iq are identical with u and v defined by (16).

From (29) and (30), λ_s and λ_a can be obtained for both vertical and horizontal polarization. The results are plotted in Fig. 8 together with experimental results.

Beat Wavelength

Because of the different wavelengths of the symmetric and antisymmetric modes, there will be a certain length of guide in which the phase difference between the two modes changes by 2π . This distance is called the beat wavelength λ_B , and is given by

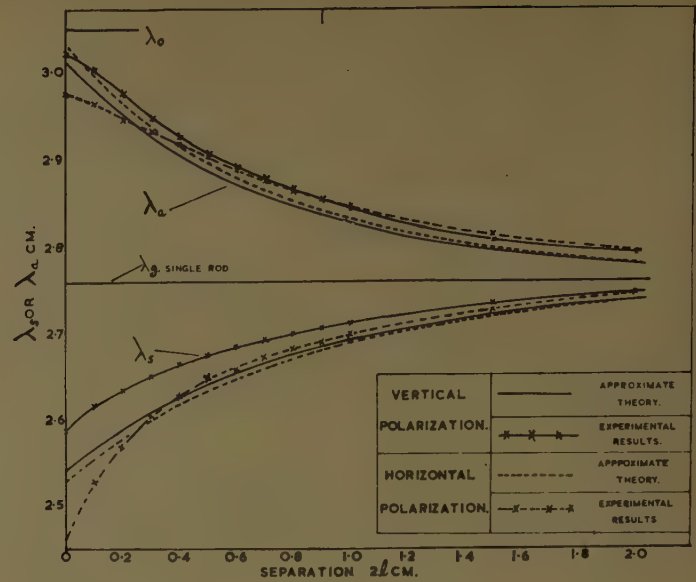


Fig. 8—Variation of λ_s and λ_a with separation at $\lambda_0=3.048$ cm. For rods 0.515 inch diameter. $\kappa=2.56$.

$$\lambda_B = \frac{\lambda_s \lambda_a}{\lambda_a - \lambda_s} \quad (33)$$

λ_B is an important parameter in the design of coupled-guide directional couplers.

In Fig. 9, theoretical and experimental results are compared both for flat sheet and rod types of guide. Agreement with experiment is good for the exact theory of coupled sheets, and reasonably good for the approximate theory of coupled rods with horizontal polarization. For vertical polarization, the theory is less satisfactory.

IV. A SURFACE-WAVE-LAUNCHING ANTENNA

Introduction

In a proposed method of studying two-dimensional diffraction problems, a directive source for surface waves was required. This ultimately took a form which can be idealized for theoretical purposes as an end-fire array of magnetic dipoles. The traveling-wave excitation has a phase velocity equal to that of the surface waves concerned.

Single Magnetic Dipole Excitation

Several authors¹¹⁻¹⁶ have dealt with the excitation of surface waves over a corrugated or dielectric-covered metal sheet, but the case of the horizontal magnetic

¹¹ A. L. Cullen, "The Excitation of Plane Surface Waves," IEE Monograph No. 93; February, 1954.

¹² D. B. Brick, "The Radiation of Hertzian Dipole over a Coated Conductor," IEE Monograph No. 113; March, 1955.

¹³ W. M. G. Fernando and H. E. M. Barlow, "An investigation of the properties of radial cylindrical surface waves launched over flat reactive surfaces," *Proc. IEE*, part B, pp. 307-318; May, 1956.

¹⁴ A. L. Cullen, "A Note on the Excitation of Surface Waves," IEE Monograph No. 239; May, 1957.

¹⁵ J. R. Wait, "Excitation of surface waves on conducting, stratified, dielectric clad and corrugated surfaces," *J. Res. Natl. Bur. Standards*, vol. 59, pp. 365-377; December, 1957.

¹⁶ B. Friedman and W. Elwyn Williams, "Excitation of Surface Waves," IEE Monograph No. 277; January, 1958.

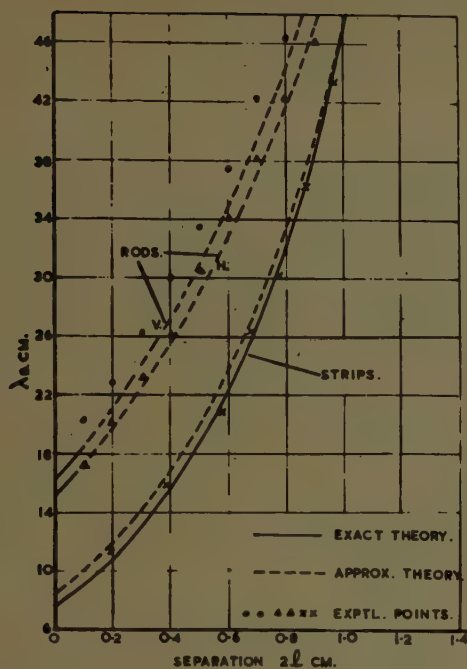


Fig. 9—Variation of beat wavelength λ_B with separation $2l$ for $\lambda_0 = 3.048$ cm. Rods 0.515 inch diameter. Strips 0.3 inch thickness. $\kappa = 2.56$.

dipole has apparently not been considered previously.

With the coordinate systems of Fig. 10, we consider a perfectly conducting plane at $z=0$, the region $0 < z < a$ being filled with dielectric of dielectric constant κ . A horizontal magnetic dipole is situated at the origin, and lies along the x -axis.

The field set up by this dipole can be split up into a part E_1, H_1 , for which $E_z=0$, and a part E_2, H_2 for which $H_z=0$. These partial fields can be expressed in terms of scalar potentials U and V through

$$E_1 = \text{curl } kV \quad (34)$$

$$H_2 = \text{curl } kU. \quad (35)$$

The potential functions can be calculated by following well-known procedures. We assume that the dielectric thickness is small enough to allow only one surface-wave mode, which arises from the U -function. The V -function contributes a radiation field only, and it is therefore convenient to express V in spherical polar coordinates. The V -function splits into a radiation term U_r , which we again express in spherical polar coordinates, and a surface-wave term U_s , which we express in cylindrical polar coordinates. The results are as follows:

$$V = -\frac{Me^{-ikr}}{4\pi r} \frac{\cos \phi \cot \theta \sqrt{\kappa - \sin^2 \theta}}{G(\sin \theta)} \quad (36)$$

$$U_r = -j^2 \frac{Me^{-ikr}}{4\pi r} \frac{\sin \phi \cot \theta}{F(\sin \theta)} \quad (37)$$

$$U_s = -\eta_0 \frac{Mk \sin \phi}{4} \frac{e^{-\sqrt{\beta_1^2 - k^2}(z-a)} H_1^{(2)}(\beta_1 \rho)}{F^1(\beta_1)}; \quad (38)$$

$(z > a)$

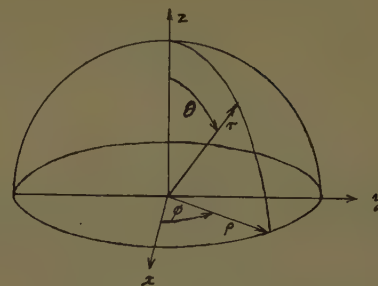


Fig. 10—Coordinate systems.

where

$$G(\sin \theta) = j \cos \theta \sin \{ka\sqrt{\kappa - \sin^2 \theta}\} + \sqrt{\kappa - \sin^2 \theta} \cos \{ka\sqrt{\kappa - \sin^2 \theta}\} \quad (39)$$

$$F(\sin \theta) = j \cos \theta \cos \{ka\sqrt{\kappa - \sin^2 \theta}\} - \frac{\sqrt{\kappa - \sin^2 \theta}}{\kappa} \sin \{ka\sqrt{\kappa - \sin^2 \theta}\} \quad (40)$$

$$F^1(\beta_1) = \frac{\beta_1(\kappa - 1)^{1/2}}{(\beta_1^2 - k^2)^{1/2}(\kappa k^2 - \beta_1^2)^{1/2}} \left[\frac{k^2}{\{(\kappa + 1)\beta_1^2 - \kappa k^2\}^{1/2}} + \frac{a\sqrt{\beta_1^2 - k^2} \{(\kappa + 1)\beta_1^2 - \kappa k^2\}^{1/2}}{\kappa} \right]. \quad (41)$$

In these equations, β_1 is the characteristic phase constant for the surface wave, k is the free-space phase coefficient, M is the magnetic current dipole moment, and $\eta_0 = (\epsilon_0/\mu_0)^{1/2}$.

Surface-Wave Array Field Patterns and Launching Efficiency

Consider now an array of n magnetic dipoles lying on the y -axis, but directed as before along the x -axis. Suppose that there is a progressive phase delay $\exp(-j\beta_1 d)$ between successive dipoles, β_1 being the surface-wave phase constant.

The V and U potentials for this system must be multiplied by an array factor as follows:

$$S = \frac{\sin \frac{n\psi}{2}}{\sin \frac{\psi}{2}}. \quad (42)$$

For the surface-wave potential U_s , ψ is given by

$$\psi = \beta_1 d (1 - \cos \phi). \quad (43)$$

For the radiation potentials U_r and V , ψ is given by

$$\psi = \beta_1 d - \beta d \sin \theta \cos \phi. \quad (44)$$

Using all these formulas, the launching efficiency, defined as $\eta = P_s/(P_r + P_s)$ in an obvious notation, has been calculated, and for a dielectric coated metal plate, $\kappa = 2.56$, $d = 0.125''$, we find $\eta = 95$ per cent at a free-space wavelength of 3.2 cm.

Diffraction by Smooth Conical Obstacles*

H. E. J. NEUGEBAUER† AND M. P. BACHYNSKI†

ABSTRACT

KIRCHHOFF'S theory for the solution of diffraction problems has been generalized to include the case of conical obstacles and energy incident at oblique incidence.

In Fig. 1, let T_1 represent a transmitter, T_2 a receiver, and Z a perfectly conducting obstacle of cylindrical shape between the two stations. Green's theorem can be used to calculate the field E at receiver T_2 .

$$E = -\frac{1}{4\pi} \iint \left(G \frac{\partial E'}{\partial n} - E' \frac{\partial G}{\partial n} \right) dS. \quad (1)$$

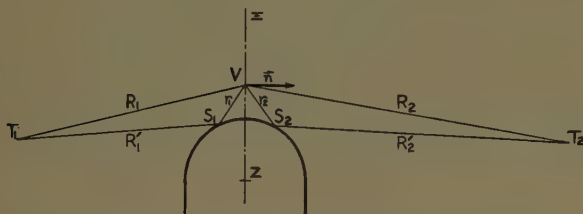


Fig. 1—Notations for diffraction by a smooth obstacle.

The integration is extended over the entire plane E separating the two stations, the normal n points inwards, E' is the radiation at the plane E and G is Green's function. If the field E' and Green's function were rigorously known, the field E could be rigorously calculated from (1). Neither E' nor G are known, but

approximate values can be obtained using the unperturbed incident field and Green's function, plus the radiation reflected on both sides of the crest of the obstacle. Since the main contribution of the reflected radiation to the integrals is supplied by rays that are almost grazing, improved values for both E' and G are obtained by using values of the electric field near the grazing point from the theory of Fock.¹ This approach is satisfactory for small values of scattering angle.

This general theory is now applied to diffraction by smooth obstacles which may be appropriately described by a cone and on which the electromagnetic radiation is incident at an angle that may be different from 90° .

Evaluation of the integrals by the stationary phase method indicates that results obtained earlier² can be applied to the halo terms in diffraction by a conical obstacle if a is replaced by $(a/\cos \tau \cos^2 \zeta)$ and $2\psi_0$ becomes $2\psi_0 \cos \tau \cos \zeta$, where a and $2\psi_0$ are the radius of curvature and scattering angle respectively, measured in the vertical plane perpendicular to the axis of the cone, and 2τ is the cone angle and ζ the angle of obliqueness (*i.e.*, 90° —angle of incidence).

Experimental measurements of power diffracted by smooth half-conical mountains have been performed using model techniques in the K -band frequency range. The measurements are found to support the theoretical predictions.

¹ V. Fock, "The distribution of currents induced by a plane wave on the surface of a conductor," *J. Phys. (USSR)*, vol. 10, pp. 130–136; February, 1946.

² H. E. J. Neugebauer and M. P. Bachynski, "Diffraction by smooth cylindrical mountains," *Proc. IRE*, vol. 46, pp. 1619–1627; September, 1958.

* This work was done under AF Cambridge Res. Ctr. Contract No. AF 19(604)-3049.

† RCA Victor Res. Labs., Montreal, Can.

Surface Waves Over a Lossy Conductor*

BERNARD FRIEDMAN†

Summary—The well-known problem of the field produced by a magnetic line dipole on the surface of a lossy conductor is investigated. Suppose ω is the frequency of the wave, μ the magnetic permeability, and σ the conductivity of the conductor. If M is the absolute value of the ratio of the actual field to the free space field, then it is found that M has a maximum value 1.55 at a distance r from the line dipole given by

$$r = 0.88\sigma(\omega^2 c \mu)^{-1}.$$

THE electromagnetic field produced by a magnetic line source on an imperfectly conducting plane has been the subject of many papers.^{1,2} The question of whether this field contains surface waves is still a matter of discussion. In this paper, we hope to avoid such an essentially semantic question as to the "existence" of a surface wave. Instead, the well-known³ integral representation for the field produced by the line source is studied, and the integrals involved are evaluated as carefully as possible. More important, estimates of errors made in evaluating the integral are obtained.

On the basis of these estimates, a simple expression for the field near the plane is obtained, which is accurate to within five per cent for such distances r of the receiver from the source that kr is greater than twelve. Here, k denotes the free-space wavelength of the radiation. The ratio of the field produced on a lossy conductor to that produced on a perfect conductor is examined, and it is found to have a maximum value of 1.55 at a distance r along the plane, such that

$$r = 0.88\sigma(\omega^2 c \mu)^{-1}$$

where ω is the circular frequency of the radiation, σ is the conductivity of the plane, μ is the magnetic permeability of the vacuum, and c is the velocity of light.

Consider a magnetic line dipole located on the y -axis at a distance h above the xz -plane, and oscillating with a time factor $e^{-i\omega t}$. The xz -plane is assumed to be an imperfect conductor, and the field at this surface is assumed to satisfy the impedance boundary condition

$$E_x/H_z = Z. \quad (1)$$

Since $E_z = H_x = H_y = 0$, the field will be completely determined by H_z . Because of the symmetry of the prob-

lem, H_z will be independent of z , and it will satisfy the equation

$$\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial y^2} + k^2 H_z = \delta(x)\delta(y-h) \quad (2)$$

where $k = \omega(\epsilon\mu)^{1/2}$.

By the use of the Fourier transform with respect to x , we easily obtain the following well-known^{2,3} representation for H_z :

$$\begin{aligned} H_z = & (4\pi i)^{-1} \int_{-\infty}^{\infty} (k^2 - p^2)^{-1/2} \\ & \cdot \exp [ipx + i(k^2 - p^2)^{1/2} |h - y|] dp \\ & + (4\pi i)^{-1} \int_{-\infty}^{\infty} (k^2 - p^2)^{-1/2} \\ & \cdot \exp [ipx + i(k^2 - p^2)^{1/2}(h + y)] \\ & \cdot \left[\frac{(k^2 - p^2)^{1/2} - \omega\epsilon Z}{(k^2 - p^2)^{1/2} + \omega\epsilon Z} \right] dp. \end{aligned} \quad (3)$$

We write

$$H_z = F_1 + F_2 \quad (4)$$

where F_1 represents the value of the first term in (3), and F_2 the value of the second term. Clearly, F_1 is the field that would be produced by the line source acting in free space, whereas F_2 is the modification produced by the lossy conductor. Note that

$$\begin{aligned} F_2 = & F_1 - \frac{2\omega\epsilon Z}{4\pi i} \int_{-\infty}^{\infty} (k^2 - p^2)^{-1/2} [(k^2 - p^2)^{1/2} + \omega\epsilon Z]^{-1} \\ & \cdot \exp [ipx + i(k^2 - p^2)^{1/2}(h + y)] dp. \end{aligned} \quad (5)$$

Put $p = kq$ in this integral and denote $\omega\epsilon Z k^{-1}$ by γ . Then (5) becomes

$$F_2 = F_1 - \frac{\gamma}{2\pi i} I \quad (6)$$

where we put $y + h = y'$ and

$$\begin{aligned} I = & \int_{-\infty}^{\infty} (1 - q^2)^{-1/2} [(1 - q^2)^{1/2} + \gamma]^{-1} \\ & \cdot \exp [ikxq + i(1 - q^2)^{1/2}ky'] dq. \end{aligned} \quad (7)$$

Combining (4) and (6), we see that

$$H_z = 2F_1 - \frac{\gamma}{2\pi i} I. \quad (8)$$

Since $\gamma = 0$ for a perfect conductor, formula (8) may be interpreted as follows:

* This research was sponsored, in part, by the Office of Naval Research, under Contract Nonr-222(60).

† University of California, Berkeley, Calif.

¹ For a discussion of the results obtained, see H. Bremmer, "Propagation of electromagnetic waves," in "Electric Fields and Waves," "Encyclopedia of Physics," Flugge, Ed., Springer, Berlin, Ger., vol. 16; 1958.

² See also J. R. Wait, "Excitation of surface waves on conducting stratified, dielectric-clad and corrugated surfaces," *J. Res. NBS*, vol. 59, pp. 365-777, December, 1957; vol. 61, pp. 205-232, September, 1958.

³ B. Friedman and W. E. Williams, "Excitation of surface waves," *Proc. IEE*, pt. C, Monograph No. 277R; January, 1958.

H_z equals twice the free-space field with a modification due to the finite conductivity of the conductor.

From the definition of γ , we have

$$\gamma = \omega\epsilon Z/k = Z(\epsilon/\mu)^{1/2} = Z/Z_0 \quad (9)$$

where Z_0 is the impedance of free space. For a lossy conductor with conductivity σ , we find

$$\gamma = (\omega\mu/\sigma)^{1/2} \exp(-i\phi) \quad (10)$$

where ϕ is slightly less than $\pi/4$.

To evaluate (7), put

$$xq + y'(1 - q^2)^{1/2} = r\xi \quad (11)$$

where

$$r^2 = x^2 + y'^2. \quad (12)$$

Note that

$$r(1 - \xi^2)^{1/2} = x(1 - q^2)^{1/2} - yq. \quad (13)$$

By differentiating (11), we find

$$[x - y'q(1 - q^2)^{-1/2}]dq = rd\xi$$

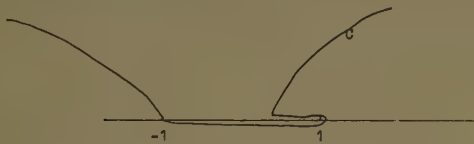
and, using (13), we get

$$(1 - q^2)^{-1/2}dq = (1 - \xi^2)^{-1/2}d\xi. \quad (14)$$

After the substitution (11), I becomes

$$r \int_C (1 - \xi^2)^{-1/2} [y'\xi + x(1 - \xi^2)^{1/2} + r\gamma]^{-1} \cdot \exp(ikr\xi) d\xi \quad (15)$$

where C is the following contour in the ξ -plane:



When we rationalize the denominator in (15), we may write

$$I = I_1 + I_2 \quad (16)$$

where

$$I_1 = -rx \int_C [(y'\xi + r\gamma)^2 - x^2(1 - \xi^2)]^{-1} \cdot \exp(ikr\xi) d\xi \quad (17)$$

and

$$I_2 = r \int_C (1 - \xi^2)^{-1/2} [(y'\xi + r\gamma)^2 - x^2(1 - \xi^2)]^{-1} \cdot (y'\xi + r\gamma) \exp(ikr\xi) d\xi. \quad (18)$$

Note that the brackets in both (17) and (18) may be written as

$$(r\xi + y'\gamma)^2 - x^2(1 - \gamma^2).$$

Integral (17) is easily evaluated by residues to give

$$I_1 = -\pi i (1 - \gamma^2)^{-1/2} \exp(ikr\xi_0) \quad (19)$$

where

$$r\xi_0 + y'\gamma = x(1 - \gamma^2)^{1/2}. \quad (20)$$

Put

$$\begin{aligned} \gamma &= \sin \psi, & (1 - \gamma^2)^{1/2} &= \cos \psi, \\ y' &= r \sin \theta, & x &= r \cos \theta, \end{aligned} \quad (21)$$

then from (20)

$$\xi_0 = \cos(\theta + \psi) \quad (22)$$

and from (19)

$$I_1 = -\pi i \sec \psi \exp[ikr \cos(\theta + \psi)]. \quad (23)$$

To evaluate (18), use (21) and note that

$$\begin{aligned} & \frac{(y'\xi + r\gamma)}{(y'\xi + r\gamma)^2 - x^2(1 - \xi^2)} \\ &= \frac{\xi \sin \theta + \sin \psi}{(\xi + \sin \theta \sin \psi)^2 - \cos^2 \theta \cos^2 \psi} \\ &= \frac{1}{2 \cos \psi} \left[\frac{\sin(\theta + \psi)}{\xi - \cos(\theta + \psi)} - \frac{\sin(\psi - \theta)}{\xi + \cos(\psi - \theta)} \right]; \end{aligned}$$

therefore

$$I_2 = (2 \cos \psi)^{-1} [I_3 \sin(\theta + \psi) - I_4 \sin(\psi - \theta)] \quad (24)$$

where

$$I_3 = \int_C (1 - \xi^2)^{-1/2} (\xi - \xi_0)^{-1} \exp(ikr\xi) d\xi \quad (25)$$

and

$$I_4 = \int_C (1 - \xi^2)^{-1/2} (\xi + \xi_1)^{-1} \exp(ikr\xi) d\xi. \quad (26)$$

Here, $\xi_1 = \cos(\psi - \theta)$.

In (25) put $\xi = 1 + i\tau$ and shift the contour so that it runs around a branch cut which goes from $\xi = 1$ to $\xi = 1 + i\infty$. Note that from (10) we find that the phase of $\xi_0 - 1$ is not less than $\pi/2$; therefore, the pole at $\xi = \xi_0$ is not crossed in this manipulation. We obtain

$$I_3 = 2 \exp(ikr - 3\pi i/4) \int_0^\infty \tau^{-1/2} (2 + i\tau)^{-1/2} (\tau - \tau_0)^{-1} \cdot \exp(-krr) d\tau \quad (27)$$

where

$$\tau_0 = -i(\xi_0 - 1) = 2i \sin^2[(\psi + \theta)/2]. \quad (28)$$

Note that the phase of τ_0 is non-negative.

For later purposes it is useful to have an estimate of the value of τ_0 . We shall always assume that ψ and θ are so small that their cosines may be approximated by one.

Then

$\sin(\psi + \theta)/2 = \sin(\psi/2) + \sin(\theta/2) = (\sin\psi + \sin\theta)/2$
approximately; consequently,

$$\tau_0 = [i(\gamma^2 + 2\gamma \sin\theta + \sin^2\theta)]/2.$$

Put $\gamma = \gamma_0 \exp(-i\pi/4)$ where γ_0 is real, and we have

$$\begin{aligned}\tau_0 &= \gamma_0^2/2 + \gamma_0 \sin\theta \exp(-i\pi/4) + (i \sin^2\theta)/2 \\ &= \alpha + i\beta\end{aligned}$$

where α and β are real. Note that

$$\begin{aligned}\alpha &= \gamma_0^2/2 + \gamma_0 \sin\theta \cos\pi/4, \\ \beta &= \sin^2\theta/2 - \gamma_0 \sin\theta \sin\pi/4.\end{aligned}\quad (29)$$

Put

$$(2 + i\tau)^{1/2} = (2 + i\tau_0)^{1/2} + i(\tau - \tau_0)R(\tau) \quad (30)$$

where

$$\begin{aligned}R(\tau) &= - (2 + i\tau)^{-1/2}(2 + i\tau_0)^{-1/2} \\ &\quad \cdot [(2 + i\tau)^{1/2} + (2 + i\tau_0)^{1/2}]^{-1}.\end{aligned}$$

For τ real between 0 and ∞ , it is easy to see that

$$|2 + i\tau| \geq 2, \quad |2 + i\tau_0| > 2,$$

and therefore

$$|R(\tau)| \leq 2^{-5/2}. \quad (31)$$

Using (30), we see that

$$I_3 = 2 \exp(ikr - 3\pi i/4) [(2 + i\tau_0)^{-1/2} I_5 + i I_6] \quad (32)$$

where

$$I_5 = \int_0^\infty \tau^{-1/2} (\tau - \tau_0)^{-1} \exp(-k\tau) d\tau \quad (33)$$

and

$$I_6 = \int_0^\infty \tau^{-1/2} \exp(-k\tau) R(\tau) d\tau. \quad (34)$$

Using (31), we find that

$$|I_6| \leq 2^{-5/2} \int_0^\infty \tau^{-1/2} \exp(-k\tau) d\tau = 2^{-5/2} (\pi/k)^{1/2}. \quad (35)$$

To evaluate I_5 we use the substitution

$$\begin{aligned}(\tau - \tau_0)^{-1} &= i \int_0^\infty \exp[-iu(\tau - \tau_0)] du \\ &= i \int_{-ikr}^{\infty - ikr} \exp[-i(u + ikr)(\tau - \tau_0)] du\end{aligned}$$

in (33). We get

$$\begin{aligned}I_5 &= i \int_{-ikr}^{\infty - ikr} \exp[i(u + ikr)\tau_0] du \int_0^\infty \tau^{-1/2} \exp(-i\tau) d\tau \\ &= (\pi)^{1/2} \exp(\pi i/4) \int_{-ikr}^{\infty - ikr} u^{-1/2} \exp(-k\tau_0 + iu\tau_0) du\end{aligned}$$

$$\begin{aligned}&= (\pi/\tau_0)^{1/2} \exp(\pi i/4 - k\tau_0) \int_{-ikr\tau_0}^{\infty - ikr\tau_0} u^{-1/2} \exp(iu) du \\ &= (\pi/\tau_0)^{1/2} \exp(\pi i/4 - k\tau_0) \\ &\quad \cdot \left[\int_{-kr\tau_0}^0 u^{-1/2} \exp(iu) du + \pi^{1/2} \exp(\pi i/4) \right] \\ &= \pi i \tau_0^{-1/2} \exp(-k\tau_0) + (\pi/\tau_0)^{1/2} \\ &\quad \cdot \exp(\pi i/4) \exp(-k\tau_0) I_7\end{aligned}\quad (36)$$

where

$$\begin{aligned}I_7 &= \int_{-ikr\tau_0}^0 u^{-1/2} \exp(iu) du \\ &= -2 \exp(-\pi i/4) \int_0^{(kr\tau_0)^{1/2}} \exp(v^2) dv.\end{aligned}\quad (37)$$

The only integral that is left to be evaluated is I_4 . Again, put $\xi = 1 + i\tau$ and shift the contour around the branch cut. We get

$$\begin{aligned}I_4 &= 2 \exp(ikr - 3\pi i/4) \int_0^\infty \tau^{-1/2} (2 + i\tau)^{-1/2} (\tau + \tau_1)^{-1} \\ &\quad \cdot \exp(-k\tau) d\tau\end{aligned}$$

where $i\tau_1 = 1 + \xi_1$. Since

$$|2 + i\tau| \geq 2 \quad \text{and} \quad |\tau + \tau_1| \geq \text{Im } \tau_1$$

we find that

$$|I_4 \exp(-ikr + 3\pi i/4)| < (2\pi/k)^{1/2} |\text{Im } \tau_1|^{-1}. \quad (38)$$

Note that

$$\begin{aligned}\exp(-3\pi i/4) \tau_0^{-1/2} (2 + i\tau_0)^{-1/2} \\ = (1 - \xi_0^2)^{-1/2} = -\csc(\psi + \theta).\end{aligned}$$

The minus sign is taken in the last equation because $\text{Im}(1 - \xi^2)^{1/2} \geq 0$. Using this equation, we find that

$$\begin{aligned}[\sin(\theta + \psi) I_3]/2 \\ = -\exp(ikr\xi_0) [\pi i + \pi^{1/2} \exp(\pi i/4) I_7].\end{aligned}\quad (39)$$

Combining this with (16), (19), and (24), we get

$$\begin{aligned}I &= -2\pi i \sec\psi \\ &\quad \cdot \exp(ikr\xi_0) - \pi^{1/2} \sec\psi \exp(ikr\xi_0 + \pi i/4) I_7 \\ &\quad - 1/2 [\sec\psi \sin(\psi - \theta) I_4].\end{aligned}\quad (40)$$

In obtaining (40), we have neglected I_6 in comparison with I_5 . That this is justified may be seen as follows:

$$|I_6/I_5| \leq 2^{-5/2} \pi^{-1/2} |kr\tau_0|^{1/2} |\exp(kr\tau_0)| (kr)^{-1}. \quad (41)$$

We shall see later that the field behavior is interesting at $|kr\tau_0| = 0.5$; therefore from (41)

$$|I_6/I_5| \leq 0.12(kr)^{-1}.$$

This shows that if $kr > 3$, then dropping I_6 makes an error of, at most, five per cent.

We shall also show that I_4 may be neglected in (40).

The ratio of the I_4 term to the first term on the right-hand side of (40) is bounded by

$$\begin{aligned} & |(4\pi)^{-1} \sin(\psi - \theta) I_4 \exp(kr\tau_0)| \\ & \leq 2^{-1}(2\pi kr)^{-1/2} |\operatorname{Im} \tau_1|^{-1} |\sin(\psi - \theta) \exp(kr\tau_0)| \quad (42) \end{aligned}$$

from (38). Now, if we assume $\sin \theta$ very much smaller than $|\sin \psi| = |\gamma|$, we find that

$$|\sin(\psi - \theta)| \leq |\gamma|$$

and then (42) is less than

$$(5kr)^{-1} |\operatorname{Im} \tau_1|^{-1} |kr\gamma^2|^{1/2} |\exp(kr\tau_0)|. \quad (43)$$

Since $|\gamma^2| = |\tau_0|$ approximately, and since $|\operatorname{Im} \tau_1| > 1$, we find that (43) is bounded by $0.6 (kr)^{-1}$. To again get an error less than five per cent, we take $kr > 12$. Our conclusion is that if $kr > 12$, then

$$\begin{aligned} I &= -2\pi i \sec \psi \\ &\cdot \exp(ikr\xi_0) [1 + (4\pi)^{-1/2} \exp(-\pi i/4) I_7] \quad (44) \end{aligned}$$

with an error of less than five per cent.

From (4), we have

$$\begin{aligned} F_1 &= (4\pi i)^{-1} \int_{-\infty}^{\infty} (k^2 - p^2)^{-1/2} \\ &\cdot \exp[ipx + i(k^2 - p^2)^{1/2} |h - y|] dp \\ &= (4\pi i)^{-1} \int_{-\infty}^{\infty} (1 - \xi^2)^{-1/2} \exp(ikr'\xi) d\xi \quad (45) \end{aligned}$$

where

$$r'^2 = x^2 + (h - y)^2.$$

With the help of the same substitution we used in (15) we get

$$\begin{aligned} F_1 &= (2\pi)^{-1} \int_0^{\infty} \tau^{-1/2} (2 + i\tau)^{-1/2} \\ &\cdot \exp(ikr' - kr'\tau - 3\pi i/4) d\tau \\ &= (2\pi)^{-1} \exp(ikr' - 3\pi i/4) \\ &\cdot \left[\int_0^{\infty} (2\tau)^{-1/2} \exp(-kr'\tau) d\tau \right. \\ &\quad \left. + \int_0^{\infty} \tau^{-1/2} R'(\tau) \exp(-kr'\tau) d\tau \right] \quad (46) \end{aligned}$$

where

$$(2 + i\tau)^{-1/2} = 2^{-1/2} + \tau^{1/2} R'(\tau). \quad (47)$$

From (47)

$$R'(\tau) = -i(2 + i\tau)^{-1/2} 2^{-1/2} [(2 + i\tau)^{1/2} + 2^{1/2}]^{-1}$$

and it is clear that

$$|R'(\tau)| < 2^{-5/2}.$$

Using this in (46), we see that

$$\begin{aligned} |F_1 - (8\pi kr')^{-1/2} \exp(ikr' - 3\pi i/4)| \\ \leq 2^{-9/2} \pi^{-1/2} (kr')^{-3/2} \end{aligned}$$

so that the per cent error in approximating F_1 by the first term of the expansion is less than

$$\frac{2^{-9/2} \pi^{-1/2} (kr')^{-3/2}}{2^{-1} (2\pi kr')^{-1/2}} = 2^{-3} (kr')^{-1}$$

and this will be less than two per cent if $kr' > 12$.

We shall assume $h=0$, in which case $r=r'$. Combining all the results, we have

$$\begin{aligned} H_z &= 2F_1 - (2\pi i)^{-1} \gamma I \\ &= (2\pi kr)^{-1/2} \exp(ikr - 3\pi i/4) \\ &\quad + \gamma \sec \psi \exp(ikr\xi_0) [1 + (4\pi)^{-1/2} \exp(-\pi i/4) I_7] \quad (48) \end{aligned}$$

with an error of less than five per cent. The first term in (48) is the field that would exist if the conductor were perfect. We call it F_0 ; then (48) may be written as follows:

$$\begin{aligned} H_z/F_0 &= 1 + \gamma(2\pi kr)^{1/2} \sec \psi \exp(-kr\tau_0 + 3\pi i/4) \\ &\cdot [1 + (4\pi)^{-1/2} \exp(-\pi i/4) I_7]. \quad (49) \end{aligned}$$

Put $\gamma = \gamma_0 \exp(-\pi i/4)$ where γ_0 is real and assume γ_0 is so small that $\sec \psi$ may be replaced by one. We get

$$\begin{aligned} H_z/F_0 &= 1 + i(2\pi kr\gamma_0^2)^{1/2} \\ &\cdot \exp(-\zeta) \left[1 + i\pi^{-1/2} \int_0^{\zeta^{1/2}} (\exp v^2) dv \right] \quad (50) \end{aligned}$$

where ζ is a nondimensional distance given by

$$\zeta = kr\tau_0.$$

The ratio P of power radiated over a lossy conductor to that radiated over a perfect conductor is

$$\begin{aligned} P &= \left[1 - (2\gamma_0^2\tau_0^{-1})^{1/2} \zeta^{1/2} \exp(-\zeta) \int_0^{\zeta^{1/2}} (\exp v^2) dv \right]^2 \\ &\quad + 2\pi\gamma_0^2\tau_0^{-1} \zeta \exp(-2\zeta). \quad (51) \end{aligned}$$

When $\theta=0$, formula (29) shows that $\tau_0 = \gamma_0^2/2$. In this case, the following table shows the values of P at different distances from the source:

| ζ | 0.2 | 0.3 | 0.4 | 0.44 | 0.5 | 0.6 | 0.7 |
|---------|------|------|------|------|------|------|------|
| P | 2.01 | 2.34 | 2.40 | 2.41 | 2.39 | 2.31 | 1.81 |

From this table, it is clear that the maximum value of P occurs for $\zeta=0.44$. This value of ζ corresponds to

$$r = 0.88\gamma_0^{-2} k^{-1} = 0.88\sigma(\omega^2 \epsilon \mu)^{-1}. \quad (52)$$

We conclude that at the distance r given by (52) the ratio of the field amplitudes has a maximum equal to $(2.41)^{1/2} = 1.55$.

It was pointed out by J. R. Wait during the discussion period that (50) is very similar to that for point source or dipole over an impedance plane.² The essential difference is that $\exp(-\zeta)$ in (51) is to be replaced by $2^{-1} \exp(-\zeta)$. In this case, the corresponding power ratio P is always less than unity for a conducting surface.

Electromagnetic Properties of Wedge and Cone Surfaces with a Linearly Varying Surface Impedance*

L. B. FELSENF

Summary—This paper presents the formal aspects of an analysis of the electromagnetic behavior of wedge and cone surfaces with a linearly varying surface impedance. Alternative field representations are obtained and their utility is described. Special attention is given to reactive surfaces which can support a new type of surface wave. This surface wave is of interest for an analysis of surface wave antennas with a linear reactance taper.

I. INTRODUCTION

WHILE the problem of diffraction of electromagnetic waves by a wedge with a constant surface impedance requires for its solution rather sophisticated mathematical techniques,¹⁻³ two-dimensional diffraction problems involving wedge and cone surfaces with a linearly varying surface impedance (or admittance, depending on the polarization of the incident field) can be solved by the conventional method of separation of variables. The comparative simplicity of the formal solutions for these variable impedance surfaces permits an explicit investigation of their guiding and scattering properties which are of interest for different applications. For example, if the surface impedance is reactive, a new type of surface wave is found to exist which can serve as a model for the analysis of a class of tapered surface-wave antennas with either rectangular (wedge problem) or circular (cone problem) symmetry. On the other hand, a study of the solutions in the quasi-optic range of short wavelengths yields the high-frequency scattering behavior of this class of variable impedance surfaces, and exhibits the dependence of the geometric-optical, diffracted, and transition range contributions to the field on the rate of variation of surface impedance. This information could be of interest for the study of diffraction of electromagnetic waves by sharp mountain ridges and peaks having certain variable surface properties, or by high-speed vehicles in the upper atmosphere surrounded by a plasma sheath with a special variable density. More important, however, is the fact that the existence of a rigorous solution for the electromagnetic behavior of certain objects with vari-

able surface properties provides a most useful means of checking approximate theories which have been proposed for the analysis of the scattering by relatively arbitrary structures.

The guiding and scattering properties of the wedge and cone surfaces with a linearly varying surface impedance, which were alluded to above, are investigated most directly through the use of alternative representations of the formal solutions which are obtained in this paper. Because of space limitations, the analysis of the tapered surface-wave antenna⁴ and quasi-optic diffraction⁵ problems will be submitted for publication in a separate paper at a future date.

II. FORMAL SOLUTIONS

A. The Wedge Configuration

1) *Formulation of the Problem:* The physical configuration and choice of cylindrical coordinates are shown in Fig. 1. The wedge is formed by the two intersecting half-

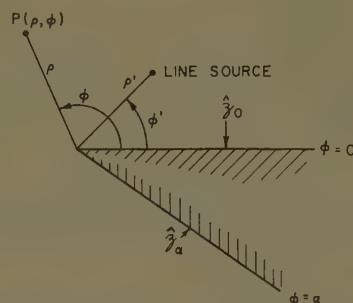


Fig. 1—Physical configuration.

planes located at $\phi = 0$ and $\phi = \alpha$, and is excited by an electromagnetic line source of constant strength extending parallel to the wedge apex and situated at $\rho' = (\rho', \phi')$. The surface impedances on the half planes at $\phi = 0$ and $\phi = \alpha$ are denoted by \hat{z}_0 and \hat{z}_α , respectively. On these surfaces, the tangential electric field \mathbf{E}_T and the tangential magnetic field \mathbf{H}_T are related by $\mathbf{E}_T = \hat{z} \mathbf{H}_T \times \mathbf{n}$, where \mathbf{n} is a unit normal vector pointing into the surface. It will be assumed that the surface impedances \hat{z}_0 and \hat{z}_α vary in such a manner as to render the associated boundary value problem separable. Consequently, we seek to infer the field solutions for various types of line

* The research reported herein was begun under a subcontract with the University of Michigan (Purchase Order 154700), and has been continued under Contract AF-19(604)-4143, sponsored by the AF Cambridge Research Center.

† Microwave Res. Inst., Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

¹ T. B. A. Senior, "Diffraction by an imperfectly conducting wedge," in *Studies in Radar Cross Section*, vol. xxv, Engrg. Res. Inst., University of Michigan, Ann Arbor, Mich.; October, 1957.

² F. C. Karal and S. N. Karp, "Diffraction of a Skew Plane Electromagnetic Wave by an Absorbing Right-Angled Wedge," *Electromagnetic Res. Div., Inst. of Math. Sci., New York University*, New York, N. Y., Res. Rept. No. EM-111; February, 1958.

³ F. C. Karal and S. N. Karp, "Diffraction of a Plane Wave by a Right-Angled Wedge which Sustains Surface Waves on One Face," *Electromagnetic Res. Div., Inst. of Math. Sci., New York University*, New York, N. Y., Res. Rept. No. EM-123; January, 1959.

⁴ The results of this analysis were included in the paper presented by the author at the Symposium on Electromagnetic Theory, University of Toronto, Toronto, Can.; June, 1959.

⁵ L. B. Felsen, "Some Aspects of Diffraction by Variable Impedance and Anisotropic Structures," *Microwave Res. Inst., Polytech. Inst. of Brooklyn, N. Y., Rep. R-685-58, PIB-613, Sec. IIIA*; September, 1958. The diffraction problem for the variable impedance wedge is discussed herein.

source excitation from the knowledge of a two-dimensional scalar Green's function $G(\varrho, \varrho')$ uniquely defined as follows:

$$\left(\frac{1}{\rho} \frac{\partial}{\partial \rho} \rho \frac{\partial}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} + k^2 \right) G(\varrho, \varrho') = -\delta(\varrho - \varrho') \quad (1)$$

with the separable boundary conditions

G non-infinite at $\rho = 0$; radiation condition at $\rho \rightarrow \infty$; (1a)

$$G = \mp c_{0,\alpha} \frac{\partial G}{\partial \phi} \text{ at } \phi = 0, \alpha, \quad c_{0,\alpha} \text{ are constants.} \quad (1b)$$

$k = \omega \sqrt{\mu \epsilon}$ is the free-space wavenumber, ω the applied angular frequency, and μ and ϵ are the permeability and dielectric constant, respectively, of the (lossless) medium. A time dependence $\exp(-i\omega t)$ is implied throughout, and the system of units is MKS.

The line source excitations considered can generally be of two types: a continuous distribution of (electric or magnetic) current elements directed either parallel or perpendicular to the apex (z axis) as shown in Fig. 2(a) and 2(b), respectively. To express the associated electro-

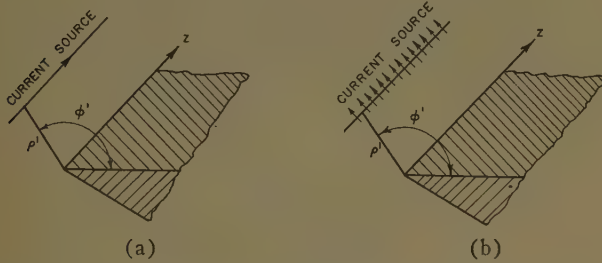


Fig. 2—Various current excitations. (a) Currents parallel to apex. (b) Currents perpendicular to apex.

magnetic fields in terms of the Green's function G we consider the Maxwell field equations with electric and magnetic current sources J and M , respectively:

$$\nabla \times E = i\omega\mu H - M, \quad \nabla \times H = -i\omega\epsilon E + J, \quad (2)$$

from which one obtains the wave equations (for $\varrho \neq \varrho'$),

$$\begin{aligned} (\nabla^2 + k^2)E &= -i\omega\mu J + \nabla \times M, \\ (\nabla^2 + k^2)H &= -i\omega\epsilon M - \nabla \times J. \end{aligned} \quad (2a)$$

For electric current excitation parallel to, or magnetic current excitation perpendicular to the apex, the total electric field has only a z component $E = z_0 E_z$, where z_0 is a unit vector along the z direction. The associated magnetic field is given by

$$H_\rho = \frac{1}{i\omega\mu\rho} \frac{\partial E_z}{\partial \phi}, \quad H_\phi = \frac{1}{-i\omega\mu} \frac{\partial E_z}{\partial \rho}. \quad (3a)$$

Upon comparing (1) and (2a) one notes that

$$E_z = i\omega\mu\delta G, \text{ when } J = z_0\delta(\varrho - \varrho'), M \equiv 0, \quad (3b)$$

while in view of $\nabla \times [\mathfrak{M}\delta(\varrho - \varrho')] = -\nabla' \times [\mathfrak{M}\delta(\varrho - \varrho')]$,

$$E_z \equiv \left[\frac{\mathfrak{M}_\phi}{\rho'} \frac{\partial}{\partial \rho'} \rho' - \frac{\mathfrak{M}_\rho}{\rho'} \frac{\partial}{\partial \phi'} \right] G,$$

when

$$M = (\varrho_0 \mathfrak{M}_\rho + \phi_0 \mathfrak{M}_\phi) \delta(\varrho - \varrho'), \quad J \equiv 0. \quad (3c)$$

ϱ_0 and ϕ_0 are unit vectors along the ρ and ϕ directions, respectively. In the radial domain, E_z is subject to the same conditions as those imposed on G in (1a). In the angular domain, $E_z = \mp z_{0,\alpha} H_\rho$ at $\phi = 0, \alpha$. To make the latter boundary condition coincide with that in (1b) one notes from (3a)–(3c) that $z_{0,\alpha}$ must have the form

$$z_{0,\alpha} = ik\rho c_{0,\alpha}, \quad z_{0,\alpha} = z_{0,\alpha} / \sqrt{\frac{\mu}{\epsilon}}, \quad (3d)$$

where $z_{0,\alpha}$ are the normalized surface impedances. Since c_0 and c_α are constant, the surface impedances, which render the problem separable, increase linearly with distance from the edge. For passive impedances, $\text{Re } z_{0,\alpha} \geq 0$; the corresponding restriction on $c_{0,\alpha}$ is $\text{Im } c_{0,\alpha} \leq 0$.

For the dual problems of magnetic current excitation parallel to, or electric current excitation perpendicular to the apex, the total magnetic field has only a single component, $H = z_0 H_z$. To express these solutions in terms of G one employs (3a)–(3d) provided that one first makes the duality replacements,

$$\begin{aligned} E &\rightarrow H, \quad H \rightarrow -E, \quad J \rightarrow M, \quad M \rightarrow -J, \\ \mu &\leftrightarrow \epsilon, \quad z_{0,\alpha} \rightarrow \frac{1}{z_{0,\alpha}}. \end{aligned} \quad (4)$$

It is to be emphasized that in this case $z_{0,\alpha} = 1/[ik\rho c_{0,\alpha}]$ so that the surface impedance decreases with distance from the edge. Thus, two different physical structures are implied in the problems stated in (3a)–(3d) and their dual as obtained from (4).

2) *Alternative Representations of the Solution:* Because of the separability of the differential operator in (1) and of the boundary conditions in (1a) and (1b) with respect to the ρ and ϕ domains, the solution for the two-dimensional Green's function G in the domain, $0 \leq \rho < \infty$, $0 \leq \phi \leq \alpha$, can be inferred from that for the two one-dimensional characteristic Green's functions G_ρ and G_ϕ defined below.^{6,7}

a) *Radial domain:* $0 \leq \rho < \infty$: The radial characteristic Green's function G_ρ is defined by the differential equation

$$\left(\frac{d}{d\rho} \rho \frac{d}{d\rho} + k^2 \rho - \frac{\lambda}{\rho} \right) G_\rho(\rho, \rho'; \lambda) = -\delta(\rho - \rho'), \quad (5)$$

subject to the boundary conditions,

G_ρ non-infinite at $\rho = 0$, radiation condition at $\rho \rightarrow \infty$, (5a)

⁶ N. Marcuvitz, "Field Representation in Spherically Stratified Regions," *Proc. New York University Symposium on the Theory of Electromagnetic Waves*, June 6–8, 1950; Interscience Publishers, Inc., New York, N. Y., 1951.

⁷ L. B. Felsen, "Alternative field representations in regions bounded by spheres, cones, and planes," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-5, pp. 109–121; January, 1957.

with the complex parameter λ so restricted as to admit a unique solution for G_ρ . The solution of (5) is given by

$$G_\rho(\rho, \rho'; \lambda) = \frac{\pi i}{2} J_\nu(k\rho_{<}) H_\nu^{(1)}(k\rho_{>}), \quad \nu = \sqrt{\lambda}, \quad (6)$$

where $\rho_{<}$ denotes ρ when $\rho < \rho'$, and ρ' when $\rho > \rho'$; the converse holds for $\rho_{>}$. Regarded as a function of λ , G_ρ is regular in the complex λ plane except for a branch point singularity at $\lambda = 0$. To assure the boundedness of G_ρ as $\rho \rightarrow 0$, the required restriction on λ is seen to be $\text{Re } \nu = \text{Re } \sqrt{\lambda} > 0$ in view of the small argument behavior of the Bessel function, $J_\nu(k\rho) \sim \rho^\nu$ as $\rho \rightarrow 0$. (This restriction also assures the vanishing of G_ρ as $|\lambda| \rightarrow \infty$ with $\text{Re } \sqrt{\lambda} > 0$.) A choice of branch cut along the negative real λ axis in the complex λ plane meets the requirement $-\pi < \arg \lambda < \pi$; i.e., $\text{Re } \sqrt{\lambda} > 0$, on the entire top sheet of the two-sheeted Riemann surface.

The formal spectral representation of the weighted delta function is inferred from a knowledge of G_ρ via the contour integration,

$$\begin{aligned} \rho' \delta(\rho - \rho') &= \frac{1}{2\pi i} \int_{C_1} G_\rho(\rho, \rho'; \lambda) d\lambda \\ &= \frac{1}{2} \int_C \nu J_\nu(k\rho_{<}) H_\nu^{(1)}(k\rho_{>}) d\nu, \end{aligned} \quad (7)$$

where the contours C_1 and C in the complex λ and ν planes, respectively, are shown in Fig. 3. Upon deform-

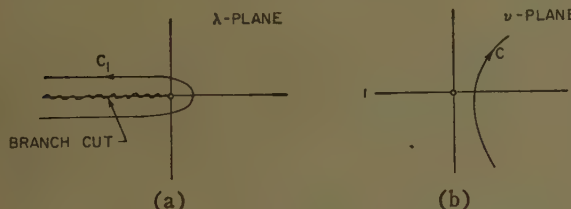


Fig. 3—Contours of integration. (a) λ plane. (b) ν plane.

ing the contour C in Fig. 3(b) into the imaginary ν axis one obtains the following formal alternative formulations

$$\rho' \delta(\rho - \rho') = \frac{1}{2} \int_{-i\infty}^{i\infty} \nu J_\nu(k\rho) H_\nu^{(1)}(k\rho') d\nu, \quad (8a)^3$$

$$= \frac{1}{4} \int_{-i\infty}^{i\infty} \nu H_\nu^{(1)}(k\rho) H_\nu^{(1)}(k\rho') d\nu, \quad (8b)$$

$$= \frac{1}{4} \int_0^{i\infty} \nu (1 - e^{i2\nu\pi}) H_\nu^{(1)}(k\rho) H_\nu^{(1)}(k\rho') d\nu. \quad (8c)$$

Eqs. (8b) and (8c) follow from (8a) upon use of the relations,

$$J_\nu(x) = \frac{1}{2} [H_\nu^{(1)}(x) + H_\nu^{(2)}(x)];$$

$$H_{-\nu}^{(1),(2)}(x) = e^{\pm i\nu\pi} H_\nu^{(1),(2)}(x). \quad (8d)$$

In applying the transform theorem implied by (8) to represent a suitable function, attention must be paid to the behavior of the integrands in (8) as $|\nu| \rightarrow \infty$. One verifies readily from the asymptotic formulas,⁹

$$J_\nu(x) \sim \left(\frac{x}{2}\right)^\nu \frac{1}{\Gamma(\nu+1)} \left[1 + O\left(\frac{1}{\nu}\right)\right], \quad |\nu| \rightarrow \infty$$

$$\Gamma(\nu + \delta) \sim \sqrt{\frac{2\pi}{\nu}} \left(\frac{\nu}{e}\right)^\nu \left[1 + O\left(\frac{1}{\nu}\right)\right], \quad |\nu| \rightarrow \infty, \quad |\arg \nu| < \pi \quad (9a)$$

and the relation,

$$H_\nu^{(1)}(y) = \frac{-e^{-i\nu\pi} J_\nu(y) + J_{-\nu}(y)}{i \sin \nu\pi}, \quad (9b)$$

that the integrands in (8) behave like $\exp[(\text{Im } \nu)\pi]$ as $\nu \rightarrow i\infty$. Thus, the transform $f(\nu)$ introduced in the integrand in the representation of a permissible function must behave at $\nu \rightarrow i\infty$ like $\exp[-(\text{Im } \nu)\varphi]$, $\varphi > \pi$, to yield a convergent representation.

b) *Angular domain:* $0 \leq \phi \leq \alpha$: The angular characteristic Green's function G_ϕ is defined by the differential equation,

$$\left(\frac{d^2}{d\phi^2} + \lambda\right) G_\phi(\phi, \phi'; \lambda) = -\delta(\phi - \phi'), \quad (10)$$

subject to the boundary conditions,

$$G_\phi = \mp c_{0,\alpha} \frac{dG_\phi}{d\phi} \text{ at } \phi = 0, \alpha, \quad (10a)$$

where c_0 and c_α are the constants defined in (3d). The complex parameter λ is so restricted that $\lambda \neq \lambda_\xi$, where λ_ξ are the eigenvalues associated with the homogeneous solutions of (10). For subsequent applications it will be convenient to view the propagation phenomena in the ϕ domain in terms of waves traveling in the $+\phi$ and $-\phi$ directions with propagation constant $\mu = \sqrt{\lambda}$; at the end-points $\phi = 0$ and $\phi = \alpha$ of the angular region, these waves are reflected with reflection coefficients Γ_0 and Γ_α , respectively, defined as

$$\begin{aligned} \Gamma_0(\mu) \equiv \Gamma_0 &= \frac{i\mu c_0 - 1}{i\mu c_0 + 1}, & \Gamma_\alpha(\mu) \equiv \Gamma_\alpha &= \frac{i\mu c_\alpha - 1}{i\mu c_\alpha + 1}, \\ \mu &= \sqrt{\lambda}. \end{aligned} \quad (11)$$

In terms of these reflection coefficients, one may write the solution for G_ϕ as follows:

$$\begin{aligned} G_\phi(\phi, \phi'; \lambda) &= \frac{(e^{-i\mu\phi} + \Gamma_0 e^{i\mu\phi})(e^{-i\mu(\alpha-\phi)} + \Gamma_\alpha e^{i\mu(\alpha-\phi)})}{-2i\mu(e^{-i\mu\alpha} - \Gamma_0 \Gamma_\alpha e^{i\mu\alpha})}. \end{aligned} \quad (12)$$

Since G_ϕ is an even function of μ , no branch point singularity exists at $\lambda = 0$ (i.e., $\sqrt{\lambda}$ need not be specially de-

³ W. Magnus and F. Oberhettinger, "Special Functions of Mathematical Physics," Chelsea Publishing Co., New York, N. Y., p. 141; 1949.

⁹ *Ibid.*, pp. 4, 17.

finer) and the only singularities of G_ϕ in the complex λ plane are simple poles situated at the zeros $\mu = \xi$, *i.e.*, $\lambda_\xi = \xi^2$, of the denominator in (12). Since $\text{Im } c_{0,\alpha} \leq 0$ [see remarks following (3d)], one can readily show (see Appendix I) that $\text{Im } \lambda_\xi \leq 0$, and that $\text{Re } \lambda_\xi$ may become negative only when $\text{Re } c_0 > 0$ or $\text{Re } c_\alpha > 0$, or both. The case, $\text{Re } \lambda_\xi < 0$, is of special interest for the propagation of surface waves along one or both of the wedge faces.

The interpretation of G_ϕ in (12) in terms of waves propagating on an infinitely extended angular transmission line emerges distinctly if one expands the denominator in (12) in terms of a series of powers of $(\Gamma_0 \Gamma_\alpha) e^{i2\mu\alpha}$:

$$G_\phi(\phi, \phi'; \lambda) = G_{\phi\infty}(\phi, \phi') + \sum_{n=1}^{\infty} \Gamma_0^{n-1} \Gamma_\alpha^n G_{\phi\infty}(\phi, 2n\alpha - \phi') \\ + \sum_{n=1}^{\infty} (\Gamma_\alpha \Gamma_0)^n [G_{\phi\infty}(\phi, 2n\alpha + \phi') + G_{\phi\infty}(\phi, -2n\alpha + \phi')] \\ + \sum_{n=0}^{\infty} \Gamma_\alpha^n \Gamma_0^{n+1} G_{\phi\infty}(\phi, -2n\alpha - \phi'), \quad \text{Im } \mu > 0, \quad (13)$$

where

$$G_{\phi\infty}(\phi, \phi') = \frac{e^{i\mu|\phi-\phi'|}}{-2i\mu}, \quad \text{Im } \mu > 0. \quad (13a)$$

The series representation in (13) converges for $\text{Im } \mu > 0$; an expansion suitable for $\text{Im } \mu < 0$ is obtained upon replacing μ everywhere in (13) and (13a) by $(-\mu)$; this procedure is valid since G_ϕ is an even function of μ . Since $G_{\phi\infty}(\phi, \phi')$ in (13a) represents the response at ϕ to a unit source situated at the point ϕ' in an infinitely extended angular transmission line (*i.e.*, $-\infty < \phi < \infty$, $-\infty < \phi' < \infty$), each term in the representation (13) can be interpreted as arising from a properly placed and weighted image source. The image sources are situated at $(2n\alpha - \phi')$, $n=0, \pm 1, \pm 2, \dots$, and $(2n\alpha + \phi')$, $n=\pm 1, \pm 2, \dots$, as shown in Fig. 4; *i.e.*, outside the



Fig. 4—Image interpretation of angular Green's function.

physical domain $0 \leq \phi \leq \alpha$. The contribution from each image source can also be identified with an angularly traveling wave in the domain $0 \leq \phi \leq \alpha$ which has experienced a given number of multiple reflections from the boundaries at $\phi = 0, \alpha$. At each reflection at $\phi = 0$ and $\phi = \alpha$, the wave amplitude changes by Γ_0 and Γ_α , respectively. This fact accounts for the different strengths of the image sources. One notes from (13) (and its appropriate form for $\text{Im } \mu < 0$) that G_ϕ behaves at $|\mu| \rightarrow \infty$, $\text{Im } \mu \neq 0$, like $e^{-|\text{Im } \mu| |\phi - \phi'|}$.

The spectral representation for the delta function $\delta(\phi - \phi')$ can now be inferred from the integral repre-

sentation^{6,10}

$$\delta(\phi - \phi') = -\frac{1}{2\pi i} \int_{\tilde{C}} G_\phi(\phi, \phi'; \lambda) d\lambda. \quad (14)$$

The contour \tilde{C} in the complex λ plane encloses all the pole singularities of G_ϕ in the positive sense. Upon evaluating the residues at these pole singularities, one obtains a series representation for $\delta(\phi - \phi')$. In order not to complicate the presentation, we assume for the present that $c_\alpha = \infty$; *i.e.*, the wedge face at $\phi = \alpha$ is perfectly reflecting ($\Gamma_\alpha = +1$). Then the poles, $\mu = \xi$ of G_ϕ in (12), are defined by the transcendental resonance equation,

$$\cot \xi \alpha = -\xi c_0. \quad (15)$$

For real values of c_0 , *i.e.*, a purely reactive surface impedance [see (3d) and (4)], the roots ξ_1, ξ_2, \dots of (15) can be obtained very simply as seen from the well-known graphical construction in Fig. 5. Of special in-

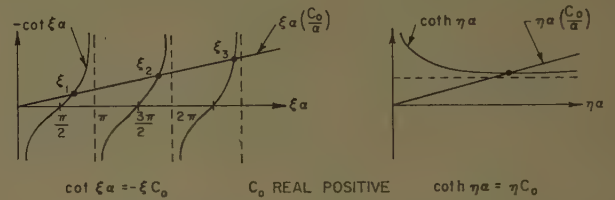


Fig. 5—Graphical solution of resonance equation.

terest is the case $c_0 > 0$, in which case an imaginary root $\xi = i\eta$, η real, exists which will be seen to correspond to a surface wave propagating along the wedge face at $\phi = 0$. Upon carrying out the residue evaluation at the poles of G_ϕ , one obtains for $c_\alpha = \infty$, c_0 real, the following representation for $\delta(\phi - \phi')$ in terms of the angular orthonormal eigenfunctions $\Phi_\xi(\phi)$:¹¹

$$\delta(\phi - \phi') = \sum_{\xi} \Phi_\xi(\phi) \Phi_\xi(\phi') + \hat{\Phi}_\eta(\phi) \hat{\Phi}_\eta(\phi') H(c_0), \quad (16)$$

where

$$\Phi_\xi(\phi) = \left[\frac{2}{\alpha \left(1 - \frac{c_0}{\alpha} \sin^2 \xi \alpha \right)} \right]^{1/2} \cos \xi(\alpha - \phi), \quad \xi > 0,$$

$$\hat{\Phi}_\eta(\phi) = \left[\frac{2}{\alpha \left(1 + \frac{c_0}{\alpha} \sinh^2 \eta \alpha \right)} \right]^{1/2} \cosh \eta(\alpha - \phi),$$

$$\eta > 0, \quad (16a)$$

¹⁰ B. Friedman, "Principles and Techniques of Applied Mathematics," John Wiley and Sons, Inc., New York, N. Y., ch. 4; 1949.

¹¹ The author is indebted to F. J. Zucker for referring him to a recent paper by V. I. Talanov, "On surface electromagnetic waves in systems with non-uniform impedance," *Izvestia VUZ MVO, Radiofizika*, vol. 2, no. 1, pp. 132-133; 1959. (This paper was translated from the Russian by Morris D. Friedman, and issued as AFCRC memo T-134, September, 1959.) Talanov has obtained independently the mode fields which can propagate along a wedge whose surface impedance varies as $1/\rho$, but has not considered the source problems (Green's functions) discussed above.

and

$$H(c_0) = \begin{cases} 1, & c_0 > 0 \\ 0, & c_0 < 0 \end{cases} \quad (16b)$$

The sum in (16) extends over all positive solutions ξ of (15). A directly analogous result is obtained for the case $c_\alpha = 0$, $c_0 > 0$, where one has the resonance relation $\tan \xi\alpha = \xi c_0$, which also admits of an imaginary root $\xi = i\eta$, $\eta > 0$, for all $0 < c_0 < \alpha$.

c) *Alternative representations for $G(\varrho, \varrho')$* : From the knowledge of the Green's function and eigenfunction solutions for the radial and angular domains, one can construct directly formal alternative representations for the two-dimensional Green's function $G(\varrho, \varrho')$ defined in (1). The most fundamental representation involves the contour integral^{6,7}

$$G(\varrho, \varrho') = -\frac{1}{2\pi i} \int_C G_\phi(\phi, \phi'; \lambda) G_\rho(\rho, \rho'; \lambda) d\lambda, \quad (17a)$$

$$= -\frac{1}{2\pi i} \int_P 2\mu G_\phi(\phi, \phi'; \mu^2) G_\rho(\rho, \rho'; \mu^2) d\mu, \quad (17b)$$

with the contour \hat{C} shown in Fig. 6(a). Upon transforming to the μ plane via $\mu = \sqrt{\lambda}$ and recalling the requirement, $\text{Re } \mu > 0$ for $G_\rho(\rho, \rho'; \lambda)$ [see (6)] one obtains the contour P shown in Fig. 6(b). The integrand in (17b) vanishes exponentially for $\varrho \neq \varrho'$ as $|\mu| \rightarrow \infty$, $|\arg \mu| < \pi/2$.

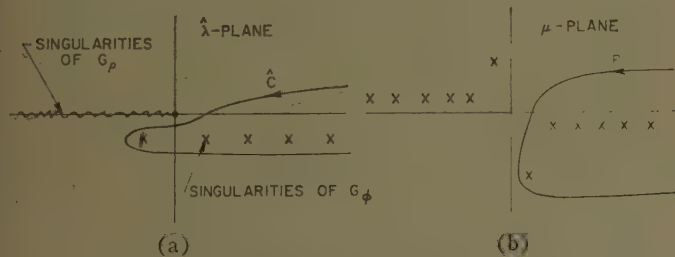


Fig. 6—Contours of integration in (a) λ and (b) μ plane ($\mu = \sqrt{\lambda}$).

If the integral in (17b) is evaluated in terms of the residues at the poles of G_ϕ , one obtains directly from (6), (14), and (16), for $c_\alpha = \infty$, $c_0 > 0$:

$$\begin{aligned} G(\varrho, \varrho') &= \frac{\pi i}{\alpha} \sum_{\xi} \frac{\cos \xi(\alpha - \phi) \cos \xi(\alpha - \phi')}{1 - \frac{c_0}{\alpha} \sin^2 \xi\alpha} J_{\xi}(k\rho_{<}) H_{\xi}^{(1)}(k\rho_{>}) \\ &+ \frac{\pi i}{\alpha} \frac{\cosh \eta(\alpha - \phi) \cosh \eta(\alpha - \phi')}{1 + \frac{c_0}{\alpha} \sinh^2 \eta\alpha} \cdot J_{-i\eta}(k\rho_{<}) H_{-i\eta}^{(1)}(k\rho_{>}), \quad (18) \end{aligned}$$

where ξ and η are defined in Fig. 5. Since $J_{\xi}(k\rho_{<})$ is very small for values of $\xi \gg k\rho_{<}$, the series representation (18) is rapidly convergent when $k\rho_{<}$ is small; i.e., when either the source point or the observation point is situated

near the wedge apex (see Appendix II for a discussion of the edge condition). When $c_0 \rightarrow \infty$, one obtains via (15): $\xi \rightarrow m\pi/\alpha$, $m = 1, 2, \dots$, $\eta \rightarrow 0$, and $\eta^2 \alpha c_0 \rightarrow 1$, $c_0 \sin^2 \xi\alpha \rightarrow 0$. Thus, one recovers the familiar formula for a wedge on whose sides $\partial G/\partial \phi$ vanishes.¹²

$$\begin{aligned} G(\varrho, \varrho') &= \frac{\pi i}{2\alpha} J_0(k\rho_{<}) H_0^{(1)}(k\rho_{>}) \\ &+ \frac{\pi i}{\alpha} \sum_{m=1}^{\infty} \cos \frac{m\pi\phi}{\alpha} \cos \frac{m\pi\phi'}{\alpha} J_{m\pi/\alpha}(k\rho_{<}) H_{m\pi/\alpha}^{(1)}(k\rho_{>}). \quad (19a) \end{aligned}$$

Similarly, for $c_0 = 0$ (i.e., $G = 0$ at $\phi = 0$, $\partial G/\partial \phi = 0$ at $\phi = \alpha$), one has $\xi = (m + 1/2)\pi/\alpha$, $m = 0, 1, 2, \dots$, $\eta \rightarrow 0$, so that

$$\begin{aligned} G(\varrho, \varrho') &= \frac{\pi i}{\alpha} \sum_{m=0}^{\infty} \sin \xi\phi \sin \xi\phi' J_{\xi}(k\rho_{<}) H_{\xi}^{(1)}(k\rho_{>}), \\ \xi &= \left(m + \frac{1}{2}\right) \frac{\pi}{\alpha}. \quad (19b) \end{aligned}$$

Of particular interest in (18) is the last term which is to be denoted $\hat{G}(\varrho, \varrho')$. If c_0 is reasonably small so that $\eta \approx 1/c_0$ [see Fig. 5(b)] and if the hyperbolic functions can be approximated in terms of the dominant exponential, one notes that

$$\begin{aligned} \hat{G}(\varrho, \varrho') &= \frac{\pi i}{c_0} e^{-\eta(\phi + \phi')} J_{-i\eta}(k\rho_{<}) H_{-i\eta}^{(1)}(k\rho_{>}), \\ \eta &\approx \frac{1}{c_0}. \quad (20) \end{aligned}$$

It is seen that the field in this mode decays exponentially away from the surface, $\phi = 0$, along any circular arc centered at the origin. Thus, it represents a new type of surface wave which can be supported on a plane surface with a linearly varying surface reactance. The conventional surface wave, which is supported on an infinite plane surface with constant surface reactance, decays exponentially away from the surface along any perpendicular plane and does not radiate any energy into the space above the surface. In contrast, the present surface wave does radiate energy; in fact, the function $\exp(-\eta\phi)$ represents essentially the radiation pattern of this wave. However, it still appears appropriate to retain the designation "surface wave" for this mode since it constitutes a proper mode of the complete spectrum of waves supported by this structure and has many of the features associated with the concept of a surface wave. It is a direct consequence of the monotonic variation of the surface reactance that this wave must radiate energy. If one considers the problem posed in (4), for example, where the magnetic field is directed parallel to the wedge apex, one notes that the normalized surface impedance $z_0 = 1/ik\rho c_0$, $c_0 > 0$, decreases with increasing distance from the apex. One may real-

¹² P. Frank and R. V. Mises, "Die Differential- und Integralgleichungen der Mechanik und Physik," Vieweg and Sons, Braunschweig, Germany; 1935. See ch. 20 by A. Sommerfeld.

ize such an impedance approximately through the use of a lossless corrugated surface, with the depth of the corrugations decreasing with increasing distance from the apex in such a manner as to yield the desired impedance variation. The energy bound to the surface is thus spread away from the surface as the surface wave propagates over increasingly shallow corrugations. Therefore, there is radiation.

Concerning further properties of the surface wave, we mention at this point the field behavior near the apex. As $\rho \rightarrow 0$, $J_{-\eta}(k\rho) \sim (k\rho)^{-i\eta} = \exp[-i\eta \ln(k\rho)]$. Since η is real for a reactive surface, this implies that the phase of the surface wave varies extremely rapidly near the edge, while its magnitude as a function of ρ remains essentially constant (note Appendix II). Moreover, it is important to point out that the surface wave can be excited in pure form by a source distribution with angular variation

$$\frac{\cosh \eta(\alpha - \phi')}{\cosh \eta\alpha}$$

on a cylindrical surface extending from $\phi=0$ to $\phi=\alpha$. This conclusion follows from the orthogonality of the surface wave mode in (18) [see also (16)] with respect to the remainder of the angular spectrum. This is again in contrast with the conventional surface wave on a constant reactive surface which requires for its pure excitation a source distribution over an infinite plane perpendicular to the surface.

It is of interest to point out that the surface wave represented in (20) is exactly that associated with a wedge whose side at $\phi=\alpha$ has a reflection coefficient $\Gamma_\alpha=0$ instead of $\Gamma_\alpha=-1$ as in (18). A reflection coefficient $\Gamma_\alpha=0$ implies that the wedge face at $\phi=\alpha$ completely absorbs all "angularly propagating" waves incident upon it. Thus, it represents a type of absorbing surface, albeit a physically unrealizable one (actually, Sommerfeld's absorber defined on an infinitely sheeted Riemann surface).¹² In terms of the infinite angular transmission-line concept introduced in (13) and the image representation of the angular characteristic Green's function, G_ϕ , one notes that $\Gamma_\alpha=0$ implies the presence of only a single image at $(-\phi')$. The associated two-dimensional Green's function comprising the contributions from the source at $\phi=\phi'$, and the image at $\phi=-\phi'$, satisfies the boundary condition in (1b) at $\phi=0$; the appropriate boundary condition at $\phi=\alpha$ is not defined except by the statement, $\Gamma_\alpha=0$, on the transform (this implies, in effect, that the ϕ domain extends to $+\infty$). It is noted that the distinction between the latter problem and that in (18) is of small consequence for any surface wave which decays with sufficient rapidity away from $\phi=0$. In fact, it would seem that the form of (20) is the simplest definition for a surface wave supported by a single surface at $\phi=0$ with a variable impedance of the type considered here. The modification, as in (18), arises from the presence of an additional physical boundary at $\phi=\alpha$. [Actually, a surface wave with ρ and ϕ de-

pendence, as in (20), can also be supported by a physical wedge configuration with $c_\alpha = -c_0 < 0$, or $\Gamma_0\Gamma_\alpha=1$. The remaining real eigenvalues are $\xi = m\pi/d$, $m=1, 2, \dots$, and the associated angular eigenfunctions are $(\xi c_0 \cos \xi\phi - \sin \xi\phi)$.

The preceding remarks concerning the case $\Gamma_\alpha=0$ are now made quantitative by constructing the corresponding two-dimensional Green's function $\bar{G}(\phi, \phi')$. We employ in this instance the contour integral representation in (17b), with G_ϕ represented as in (13).¹³ Upon extracting the residue at the pole $\xi = -i/c_0 \equiv -i\eta$ of Γ_0 , one can deform the contour P into the two contours P_1+P_2 , where P_1 runs from ∞ to 0 slightly above the real μ axis, while P_2 runs from 0 to ∞ slightly below the real μ axis. Upon inserting the appropriate representations for G_ϕ from (13) one obtains:

$$\begin{aligned} \bar{G}(\phi, \phi') = & 2\eta e^{-\eta(\phi+\phi')} g_\rho(\rho, \rho'; -i\eta) \\ & + \frac{1}{2\pi} \int_0^\infty \left[e^{i\mu|\phi-\phi'|} + \frac{\mu+i\eta}{\mu-i\eta} e^{i\mu(\phi+\phi')} \right. \\ & \left. + e^{-i\mu|\phi-\phi'|} + \frac{\mu-i\eta}{\mu+i\eta} e^{-i\mu(\phi+\phi')} \right] g_\rho(\rho, \rho'; \mu) d\mu, \quad (21) \end{aligned}$$

where $\eta = 1/c_0$ and

$$g_\rho(\rho, \rho'; \mu) = \frac{\pi i}{2} J_\mu(k\rho<) H_\mu^{(1)}(k\rho>). \quad (21a)$$

The first term in (21) is identical with \hat{G} in (20). The remainder of the ϕ -spectrum is now continuous since the ϕ -domain extends from $\phi=0$ to $\phi=\infty$. One verifies quite simply from (21) that the surface wave term is orthogonal to the continuous spectrum as represented by the integral term. Upon multiplying (21) by the surface wave mode function $e^{-\eta\phi}$, integrating over ϕ from $\phi=0$ to $\phi=\infty$, interchanging the orders of the ϕ and μ integrations and evaluating the trivial ϕ integrals, one finds that the resulting integrand vanishes for all μ [note: in the integrand, $|\phi-\phi'|$ can be replaced by $(\phi-\phi')$, because of the occurrence of $\exp(i\mu|\phi-\phi'|) + \exp(-i\mu|\phi-\phi'|)$]. A Green's function analogous to \bar{G} in (21) has been employed in the analysis of radiation from a tapered surface wave antenna.⁴

A representation for $G(\phi, \phi')$ suitable for an evaluation of the quasi-optic effects when both $k\rho$ and $k\rho'$ are large is obtained upon deforming the contour P in (17b)

¹³ The Green's function $G_{\phi\infty}(\phi, \phi')$ satisfying (10) in the domain $-\infty < (\phi, \phi') < \infty$ and representing outgoing waves at $\phi \rightarrow \pm\infty$ is given in the range $-\pi < \arg \hat{\lambda} < \pi$, i.e., $\text{Re} \sqrt{\hat{\lambda}} > 0$, by [see (13a)]

$$G_{\phi\infty}(\phi, \phi') = \frac{e^{\pm i\mu|\phi-\phi'|}}{\mp 2i\mu}, \quad \text{Im } \mu \geq 0, \quad \text{Re } \mu > 0, \quad \mu = \sqrt{\hat{\lambda}}.$$

Thus, there exists a branch point singularity at $\hat{\lambda}=0$, and $G_{\phi\infty}$ is discontinuous across the positive real μ -axis. The corresponding contour P in Fig. 6(b) must therefore enclose the entire positive real μ -axis; i.e., it must pass through the origin $\mu=0$. For the angular Green's function,

$$\bar{G}_\phi(\phi, \phi) = G_{\phi\infty}(\phi, \phi') + \Gamma_0 G_{\phi\infty}(\phi, -\phi'),$$

satisfying the boundary condition in (1b) at $\phi=0$, the contour P encloses the positive real μ -axis and the possible pole at $\mu = -i/c_0$ when $\text{Re } c_0 > 0$.

into the imaginary μ axis. This constitutes a representation in terms of the radial eigenfunctions as in (8a)–(8c). For a study of the diffracting properties of the wedge, the formulation in (8c) is particularly useful and yields⁵

$$G(\vartheta, \vartheta') = \frac{1}{4} \int_0^{i\infty} \mu (1 - e^{i2\mu\pi}) H_{\mu}^{(1)}(k\rho) H_{\mu}^{(1)}(k\rho') G_{\phi}(\phi, \phi'; \mu^2) d\mu, \quad |\phi - \phi'| > \pi. \quad (22)$$

Although the expression in (17b) is valid for all $\vartheta \neq \vartheta'$ provided that the endpoints of the contour P lie at $|\mu| \rightarrow \infty$, $|\arg \mu| < \pi/2$, the deformation of P into the imaginary axis ($|\arg \mu| = \pi/2$) requires the imposition of certain restrictions. As noted from (9a) and (9b), $H_{\mu}^{(1)}(k\rho) H_{\mu}^{(1)}(k\rho')$ behaves like $\exp(|\mu|\pi)$ as $\mu \rightarrow i\infty$. Since G_{ϕ} behaves like $\exp(-|\mu| |\phi - \phi'|)$ as $\mu \rightarrow i\infty$ [see (13)], one requires for the existence of the integral in (22) the restriction $|\phi - \phi'| > \pi$. The angular domain $|\phi - \phi'| > \pi$ is seen from Fig. 1 to correspond to the geometrical shadow region; *i.e.*, the representation in (22) is valid only for those observation points from which the source is not directly visible. This implies for a quasi-optic evaluation that $G(\vartheta, \vartheta')$, as expressed in (22), represents only a diffracted, and not a geometric-optical, contribution. This fact can also be appreciated from the form of the integrand which contains a radial dependence in terms of $H_{\nu}^{(1)}$ functions only, *i.e.*, outgoing waves. Thus, the integral certainly cannot represent the incident wave contribution.

B. The Cone Configuration

1) *Formulation and Solution of the Problem:* The physical configuration and choice of spherical coordinates are shown in Fig. 7. The semi-infinite cone is defined by $\theta = \theta_0$ and it is assumed that it is excited by a ring source of constant strength centered on the cone axis and defined by the coordinates (r', θ') . (In view of the symmetry of the source and the conical obstacle with respect to the azimuthal ϕ variable, the formulation of the problem is independent of ϕ .) As for the wedge problem, it is assumed that the surface impedance z on the cone varies with r in such a manner as to render the resulting boundary value problem separable. The ring source distributions to be treated are of two types: azimuthal or radial electric or magnetic currents as shown in Fig. 8(a) and 8(b), flowing along the ϕ and r directions, respectively. For the radial magnetic and azimuthal electric current excitations, the total electric field has only a single component E_{ϕ} in terms of which the magnetic field components H_r and H_{θ} are given by

$$H_r = \frac{1}{i\omega\mu r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta E_{\phi}),$$

$$H_{\theta} = \frac{-1}{i\omega\mu r} \frac{\partial}{\partial r} (r E_{\phi}). \quad (22a)$$

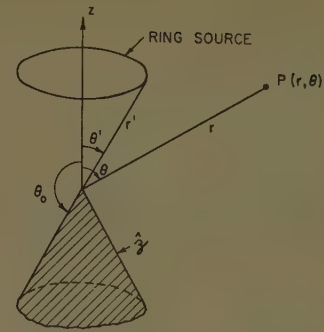


Fig. 7—Cone with ring source excitation.

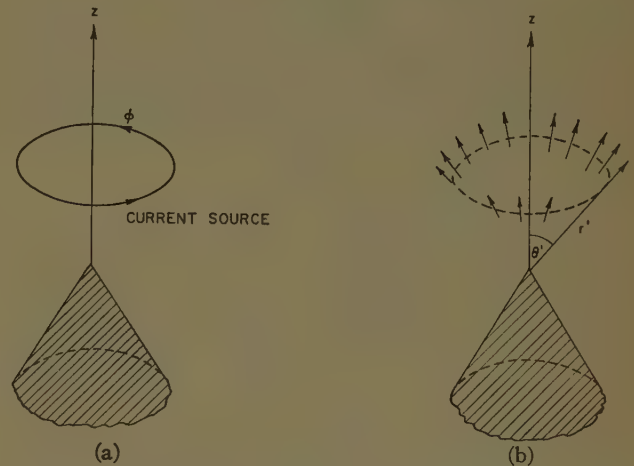


Fig. 8—Various ring source excitations. (a) Azimuthal currents. (b) Radial currents.

Conversely, for the radial electric and azimuthal magnetic current excitations, the total magnetic field has only a single component H_{ϕ} , in terms of which E_r and E_{θ} are given by

$$E_r = \frac{-1}{i\omega\epsilon r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta H_{\phi}),$$

$$E_{\theta} = \frac{1}{i\omega\epsilon r} \frac{\partial}{\partial r} (r H_{\phi}). \quad (22b)$$

The solutions for E_{ϕ} and H_{ϕ} in (22a) and (22b), respectively, can be expressed in terms of auxiliary scalar functions. Since the perfectly reflecting ring-source excited cone has been investigated in a previous publication, the results obtained there can form the starting point for the present treatment in which a variable surface impedance is assumed. For the case of a ring source of radial magnetic currents, *i.e.*, $\mathbf{M}(\mathbf{r}) = r_0 \mathbf{u}_r \delta(r-r') \delta(\theta-\theta')/r'$, where r_0 is the radial unit vector, the azimuthal electrical field E_{ϕ} can be derived from a Green's function $G(\mathbf{r}, \mathbf{r}')$ via the relation,¹⁴

$$E_{\phi} = \frac{\mathfrak{M}}{r'} \frac{\partial G}{\partial \theta}. \quad (23)$$

¹⁴ L. B. Felsen, "Radiation from ring sources in the presence of a semi-infinite cone," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-7, pp. 168–180; April, 1959. See (34b).

A contour integral representation for G analogous to that in (17b) is given by¹⁴

$$G(r, r') = -\frac{1}{2} \sqrt{\frac{r'}{r}} \sin \theta' \cdot \int_P \mu J_\mu(kr_<) H_\mu^{(1)}(kr_>) G_\theta(\theta, \theta'; \mu) d\mu, \quad (24)$$

where the angular Green's function, G_θ , satisfies the differential equation

$$\left[\frac{d}{d\theta} \sin \theta \frac{d}{d\theta} + \left(\mu^2 - \frac{1}{4} \right) \sin \theta \right] G_\theta = -\delta(\theta - \theta'). \quad (24a)$$

The contour P shown in Fig. 9 runs along a semicircle at infinity in the right-half μ plane and along the imaginary μ axis.

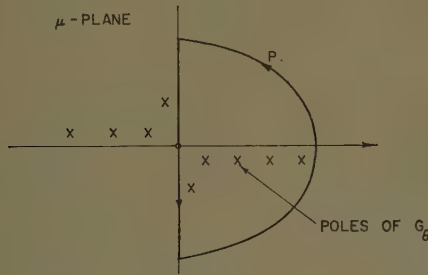


Fig. 9—Contour of integration in μ plane.

It was shown previously for the perfectly reflecting case ($E_\phi = 0$; i.e., $\partial G / \partial \theta = 0$ at $\theta = \theta_0$) that the representation in (24) constitutes the solution for the ring source problem involving radial magnetic currents provided that $\partial G_\theta / \partial \theta = 0$ at $\theta = \theta_0$. For the present problem the boundary condition on the cone surface is, instead,

$$E_\phi = \pm H_r \text{ at } \theta = \theta_0, \quad (25a)$$

which, from (22a) and (23), leads to the following condition on G :

$$\frac{\partial G}{\partial \theta} = c \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial G}{\partial \theta}, \quad c = \frac{z}{ikr}, \quad z = \pm \sqrt{\frac{\epsilon}{\mu}}. \quad (25b)$$

One notes from (25b) and (24) that the required boundary condition on G can be satisfied in terms of G_θ alone if $c = \text{constant}$. This implies that the surface impedance variation appropriate to a treatment by separation-of-variables techniques is $z(r) = ickr$, $c = \text{constant}$. Making this assumption, one has from (25b), (24), and (24a),

$$\frac{dG_\theta}{d\theta} = -c(\mu^2 - \frac{1}{4})G_\theta \text{ at } \theta = \theta_0, \quad c = \text{constant}. \quad (25c)$$

One readily constructs the solution for G_θ in (24a) subject to the boundary condition (25c) at $\theta = \theta_0$, and to a requirement of finiteness at the singular endpoint, $\theta = 0$:

$$G_\theta(\theta, \theta'; \mu) = -\frac{\pi}{2} \frac{P_\nu(\cos \theta_<)[P_\nu(-\cos \theta_>) + A_\nu P_\nu(\cos \theta_>)]}{\sin \nu \pi},$$

$$\nu = \mu - \frac{1}{2}, \quad (26)$$

where

$$A_\nu = \frac{P_\nu^1(-\cos \theta_0) - c(\nu + 1)\nu P_\nu(-\cos \theta_0)}{P_\nu^1(\cos \theta_0) + c(\nu + 1)\nu P_\nu(\cos \theta_0)},$$

$$P_\nu^1(\cos \theta) = \frac{d}{d\theta} P_\nu(\cos \theta), \quad (26a)$$

and $P_\nu^m(\cos \theta)$ is the associated Legendre function of degree ν , order m and argument $\cos \theta$.

$$\left[\text{Note: } P_\nu^1(-\cos \theta_0) = -\frac{d}{d\theta_0} P_\nu(-\cos \theta_0). \right]$$

As before, $\theta_>$ denotes θ when $\theta > \theta'$, and θ' when $\theta < \theta'$; the converse holds for $\theta_<$. Since $P_\nu^m = P_{-\nu-1}^m$, G_θ is an even function of μ . From the asymptotic formula¹⁵

$$P_\nu^m(\cos \theta) \sim \nu^m \sqrt{\frac{2}{\pi \nu \sin \theta}} \cos \left[\left(\nu + \frac{1}{2} \right) \theta + \frac{m\pi}{2} - \frac{\pi}{4} \right], \quad (26b)$$

valid when $|\nu| \rightarrow \infty$, $|\arg \nu| < \pi$, $\sin \theta \neq 0$, one notes that

$$|G_\theta| \sim \frac{\text{constant}}{|\mu| \sqrt{\sin \theta \sin \theta'}} \exp - [|\operatorname{Im} \mu| |\theta - \theta'|],$$

$$|\mu| \rightarrow \infty, \quad 0 < |\arg \mu| < \pi. \quad (26c)$$

Since $P_\nu(1) = 1$, (26c) can also be employed for $\theta' = 0$ or $\theta = 0$, provided that $\sin \theta'$ or $\sin \theta$, respectively, is replaced by unity.

For the case of a ring source distribution of azimuthal magnetic currents in the presence of a perfectly reflecting semi-infinite cone,¹⁶ i.e., $\mathbf{M}(r) = \phi_0 \mathcal{M} \delta(r - r') \delta(\theta - \theta') / r'$, where ϕ_0 is a unit vector in the ϕ direction, the magnetic field $\mathbf{H}(r) \equiv \phi_0 H_\phi(r)$ can be inferred from a scalar function $S(r, r')$ via¹⁶

$$H_\phi = \frac{i\omega \epsilon \mathcal{M}}{r} \sin \theta' \frac{\partial^2}{\partial \theta \partial \theta'} S(r, r'), \quad (27)$$

where

$$S(r, r') = -\frac{1}{2} \sqrt{rr'} \int_P \frac{\mu}{\mu^2 - \frac{1}{4}} J_\mu(kr_<) H_\mu^{(1)}(kr_>) G_\theta(\theta, \theta'; \mu) d\mu. \quad (27a)$$

The contour P in the complex μ plane is shown in Fig. 9. G_θ satisfies the differential equation (24a), and since $E_r = 0$ at $\theta = \theta_0$ for the perfectly conducting cone, one finds from (27), (27a), and (22b) that, correspondingly,

¹⁵ *Op. cit.*, W. Magnus and F. Oberhettinger, p. 71.

¹⁶ *Op. cit.*, L. B. Felsen, (38a) and (38b).

$G_\theta = 0$ at $\theta = \theta_0$. Upon comparing (24) and (27a), one notes in view of (24a) that G can be obtained from S through differentiation:

$$G = -\frac{\sin \theta'}{r} \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial S}{\partial \theta}, \quad \theta \neq \theta'. \quad (27b)$$

If the cone has a surface impedance \hat{z} , then

$$E_r = -\hat{z}H_\phi \text{ at } \theta = \theta_0, \quad (28a)$$

leading from (27), (27a), and (22b) to the boundary condition in (25c) provided that the constant c is defined as

$$c = \frac{1}{ikrz} = \text{constant}. \quad (28b)$$

Thus, for the present case, the impedance variation is $z(r) = (ikrc)^{-1}$, in contrast to (25b). This behavior is directly analogous to that encountered in (3d) and (4). The solution for G_θ is again given by (26), with c defined as in (28b).

The solutions for the remaining problems involving ring-source excitations with radial or azimuthal electric currents are obtained from the above after making the usual duality replacements as noted in (4).

2) *Alternative Representations*: From the preceding discussion, it is noted that the solutions for the various types of ring source excitation can be obtained from the functions G or S defined in (24) and (27a), respectively. Instead of the contour integrals taken over the path P in Fig. 9, one may obtain a series representation involving the angular eigenfunctions by deforming the contour about the singularities of G_θ . These singularities are the poles $\mu = \xi$ of A_ν defined in (26a):

$$B_\xi(\theta_0) = 0, \quad \text{Re } \xi \geq 0,$$

$$B_\mu(\theta_0) = P_{\mu-1/2}^1(\cos \theta_0) + c(\mu^2 - \frac{1}{4})P_{\mu-1/2}(\cos \theta_0). \quad (29)$$

No poles exist at the zeros $\nu = m$, $m = 1, 2, \dots$, of $\sin \nu\pi$ since $P_m(-x) = (-1)^m P_m(x)$, so that the numerator in (26) also vanishes. G_θ does have a pole at $\nu = 0$. [Note

$\mu = \frac{1}{2}$. Thus, the pole at $\mu = \frac{1}{2}$ does not contribute to the electromagnetic fields and will henceforth be ignored.

For a passive surface impedance, $\text{Re } z \geq 0$, and $\text{Im } c \leq 0$, as seen from (25b) and (28b). One can then show (see Appendix I) that $\text{Im } \xi \leq 0$ in (29). For simplicity, we examine only the case c real, appropriate to a reactive impedance. Eq. (29) then has an infinity of real solutions whose approximate values for large positive ξ can be inferred from the asymptotic representation in (26b):

$$\tan(\xi\theta_0 - \pi/4) \approx c\xi, \quad \xi \sin \theta_0 \gg 1. \quad (30)$$

This equation can be plotted graphically in analogy with Fig. 5. Of special interest is the case $c > 0$, for which (29) admits of an imaginary solution $\xi = \pm i\eta$, $\eta > 0$.¹⁷ For imaginary ξ , (29) takes on the following form,

$$K_\eta^1(\cos \theta_0) = c(\eta^2 + \frac{1}{4})K_\eta(\cos \theta_0), \\ K_\eta^m \equiv P_{m-1/2+i\eta}^m \equiv K_{-m}^m. \quad (31)$$

Since the conical functions $K_\eta(\cos \theta_0)$ and $K_\eta^1(\cos \theta_0)$ are positive functions when η is real and $0 \leq \theta_0 \leq \pi$,¹⁸ it is evident that (31) has a real solution only when $c > 0$. If it is assumed that η is to be so large that the asymptotic formula,

$$K_\eta(\cos \theta_0) \sim \frac{1}{\sqrt{2\pi\eta \sin \theta_0}} e^{\eta\theta_0}, \quad \eta \sin \theta_0 \gg 1, \quad (32)$$

can be applied, then η is given approximately by

$$\eta \approx \frac{1}{c}, \quad \eta \sin \theta_0 \gg 1. \quad (33)$$

The analogy between this result and that in (20) is to be noted.

Upon carrying out an evaluation of the integral in (27a) in terms of the residues at the poles $\mu = \xi$, $\xi > 0$ and $\mu = -i\eta$, $\eta > 0$, of the integrand, one obtains for $c > 0$ the following representation in terms of the angular eigenfunctions (the irrelevant pole at $\mu = \frac{1}{2}$ has been ignored):

$$S(r, r') = -\pi i \sqrt{rr'} \sum_{\xi} \frac{\xi}{\xi^2 - \frac{1}{4}} J_\xi(kr<) H_\xi^{(1)}(kr>) \frac{P_{\xi-1/2}(\cos \theta) P_{\xi-1/2}(\cos \theta')}{P_{\xi-1/2}(\cos \theta_0) \left[\frac{\partial}{\partial \mu} B_\mu(\theta_0) \right]_{\mu=\xi}} \frac{1}{\sin \theta_0} \\ - \frac{\pi i \sqrt{rr'} \eta}{\sin \theta_0 (\eta^2 + \frac{1}{4})} J_{-i\eta}(kr<) H_{-i\eta}^{(1)}(kr>) \frac{K_\eta(\cos \theta) K_\eta(\cos \theta')}{K_\eta(\cos \theta_0) \left[\frac{\partial}{\partial x} \{ K_x^1(\cos \theta_0) - c(x^2 + \frac{1}{4}) K_x(\cos \theta_0) \} \right]_{x=\eta}}. \quad (34)$$

$P_0(x) = 1$, $P_\nu^1(x) \propto \nu$ as $\nu \rightarrow 0$], and, consequently, the integrands in the expression for G in (24) and S in (27a) have simple and double poles, respectively, at $\nu = 0$; i.e., $\mu = \frac{1}{2}$. However, for the evaluation of the electromagnetic fields, the pertinent quantities are $\partial G / \partial \theta$ in (23) and $\partial^2 S / \partial \theta \partial \theta'$ in (27). Since $(d/d\theta)P_\nu(\pm \cos \theta) \propto \nu$ as $\nu \rightarrow 0$, it follows that the integrands which result upon carrying out the derivative operations are regular at

$B_\mu(\theta_0)$ is defined in (29). In obtaining this form of the result, it has been found useful to employ the Wronskian⁶

¹⁷ The special case $c > 0$ is deduced from the general case of complex c by letting $\text{Im } c \rightarrow 0$. Since all pertinent poles of G_θ in the right half of the μ plane are located in the fourth quadrant when c is complex, one notes that only the imaginary pole at $\mu = -i\eta$, $\eta > 0$ is relevant for $c > 0$.

¹⁸ *Op. cit.*, P. Frank and R. V. Mises, p. 74.

$$P_\nu(\cos \theta) P_\nu^1(-\cos \theta) + P_\nu(-\cos \theta) P_\nu^1(\cos \theta) \\ = -\frac{2 \sin \nu \pi}{\pi \sin \theta_0} \quad (34a)$$

Like (18), this series representation is rapidly convergent when either kr or kr' is small (see Appendix II for a discussion of the "tip condition" for the cone).

As in the analogous representation for the wedge problem in (18), the last term in (34) represents a surface wave which has many of the same characteristics as that in (18). To highlight these similarities we examine this last term in the range of large η , for which $\eta \approx 1/c$, c small [see (33)]. One may then employ the asymptotic form (32) for the conical functions. Use of (27) then yields the following surface wave contribution to the magnetic field (to be denoted by \hat{H}_ϕ) caused by a ring source of azimuthal magnetic currents:

$$\hat{H}_\phi(r, r') \approx \sqrt{\frac{r' \sin \theta'}{r \sin \theta}} \\ \cdot [-\pi \omega \epsilon \mathcal{M} \eta e^{-\eta(2\theta_0 - \theta - \theta')} J_{-i\eta}(kr_<) H_{-i\eta}^{(1)}(kr_>)], \quad (35)$$

where η is large and $\eta \approx 1/c$. The expression inside the square brackets in (35) is identical with the surface-wave contribution to the magnetic field, *i.e.*, $\hat{H}_z = i\omega \epsilon \mathcal{M} \hat{G}$, calculated from (20), (3b), and (4) for the wedge problem excited by a z -directed line source of magnetic currents, provided that ρ, ρ', ϕ, ϕ' for the wedge are replaced by $r, r', (\theta_0 - \theta), (\theta_0 - \theta')$ for the cone (the dependence on $(\theta_0 - \theta)$ and $(\theta_0 - \theta')$; *i.e.*, the actual angles of elevation above the cone surface, arises because the cone surface is defined by $\theta = \theta_0$, whereas the appropriate wedge surface in (20) is that at $\phi = 0$). The factor $(r' \sin \theta' / r \sin \theta)^{1/2}$ in (35) accounts for the spreading which takes place in a spherical wave.

A significant difference between the angular spectrum in (18) and that in (34) is that the latter does not satisfy the conventional orthogonality relation. This property is caused by the boundary condition in (25c) which involves the eigenvalue $\mu = \xi$. An appropriate modified orthogonality condition is given in Appendix I.

As for the wedge problem (13), it is possible to represent the cone Green's function in terms of images situated at appropriate points in an infinitely extended θ -transmission line. Such a representation has been obtained for a perfectly reflecting cone in (17) of Felsen.¹⁴ We discuss here only the case which is directly analogous to that in (21); *i.e.*, the contributions from the source at $\theta = \theta'$ and the image at $\theta = 2\theta_0 - \theta'$ alone are taken into account (see Fig. 4, with $\alpha \equiv \theta_0, \phi \equiv \theta, \phi' \equiv \theta'$). The resulting Green's function satisfies the variable impedance boundary condition (25b) at θ_0 ; at $\theta = 0$, on the other hand, the termination is one which perfectly absorbs all "angularly propagating waves," *i.e.*, the $\theta = 0$ axis is surrounded by a "black cone," the counterpart in spherical geometry of Sommerfeld's "black screen."¹² The latter condition implies, in effect, that the angular

domain extends from $-\infty < \theta \leq \theta_0$. The resulting surface wave supported by the variable impedance cone surface at $\theta = \theta_0$ is simpler in form than that in (34), and forms a useful model for the analysis of radiation from a tapered surface wave antenna with azimuthal symmetry.

The angular Green's function $\bar{G}_\theta(\theta, \theta'; \mu)$, appropriate to an infinitely extended θ domain, comprises only waves traveling away from the source and can be represented as¹⁹

$$\bar{G}_\theta(\theta, \theta'; \mu) = \frac{\pi i}{4} P_\mu^{(2)}(\theta_<) P_\mu^{(1)}(\theta_>), \quad (36)$$

where the traveling wave functions $P_\mu^{(1),(2)}$ are defined in terms of the standing wave functions $P_{\mu-1/2}(\pm \cos \theta)$ as

$$P_\mu^{(1),(2)}(\theta) = \mp \frac{i}{\cos \mu \pi} \cdot [P_{\mu-1/2}(-\cos \theta) - e^{\pm i(\mu-1/2)\pi} P_{\mu-1/2}(\cos \theta)]. \quad (36a)$$

Alternatively, one may employ a representation in terms of the hypergeometric function $F(a, b; c; z)$,

$$P_\mu^{(1),(2)}(\theta) = \sqrt{\frac{2}{\pi \sin \theta}} \frac{\Gamma(\mu + \frac{1}{2})}{\Gamma(\mu + 1)} \cdot e^{\pm i(\mu - \pi/4)} R_\mu^{(1),(2)}(\theta), \quad (36b)$$

$$R_\mu^{(1),(2)}(\theta) = (1 - e^{\pm i2\theta})^{1/2} F(\frac{1}{2}, \mu + \frac{1}{2}; \mu + 1; e^{\pm i2\theta}), \quad (36c)$$

where $\Gamma(z)$ is the gamma function. The expressions in (36b) and (36c) are particularly convenient for extending the range of θ from the restricted domain $0 < \theta < \pi$, for which the right-hand side of (36a) also applies, to the infinite domain $-\infty < \theta < \infty$. Thus, for integral n ,

$$P_\mu^{(1),(2)}(\theta + n\pi) = e^{\pm in\mu\pi} P_\mu^{(1),(2)}(\theta). \quad (36d)$$

One notes from (36b) and (36c) that the Green's function in (36) is subject to the restriction $\text{Im } \mu > 0$. An alternative representation valid for $\text{Im } \mu < 0$ is given by $\bar{G}_\theta(\theta, \theta'; -\mu)$.

To satisfy the boundary condition (25c) at $\theta = \theta_0$, we employ the representation (for $\text{Im } \mu > 0$)

$$\bar{G}_\theta(\theta, \theta'; \mu) = \frac{\pi i}{4} P_\mu^{(2)}(\theta_<) [P_\mu^{(1)}(\theta_>) + \hat{\Gamma} P_\mu^{(2)}(\theta_>)], \quad (37)$$

and determine the reflection coefficient $\hat{\Gamma}$ as

$$\hat{\Gamma} = -\frac{\frac{d}{d\theta_0} P_\mu^{(1)}(\theta_0) + c(\mu^2 - \frac{1}{4}) P_\mu^{(1)}(\theta_0)}{\frac{d}{d\theta_0} P_\mu^{(2)}(\theta_0) + c(\mu^2 - \frac{1}{4}) P_\mu^{(2)}(\theta_0)}. \quad (37a)$$

¹⁹ *Op. cit.*, L. B. Felsen, (14)–(17) and Appendix I. An alternative formulation for \bar{G}_θ has recently been employed in a somewhat different context by P. C. Clemmow, "An Infinite Legendre Integral Transform and Its Inverse," Studies in Radar Sections XXXIV, Rad. Lab., University of Michigan, Ann Arbor, Rep. no. 2778-5-T, Appendix, March, 1959.

The function \bar{S} appropriate to the present problem is then obtained upon replacing G_θ in (27a) by \bar{G}_θ , where it is understood that $\bar{G}_\theta(\theta, \theta'; \mu)$ is to be employed on those portions of the path P (Fig. 9) on which $\text{Im } \mu > 0$, while $\bar{G}_\theta(\theta, \theta'; -\mu)$ is to be used for $\text{Im } \mu < 0$. Since the only singularities of $P_\mu^{(1),(2)}$ in the complex μ plane are

$$\begin{aligned} \bar{H}_\phi(r, r') = & -\frac{\pi^2 i \omega \epsilon \mathcal{M} \sin \theta'}{4\zeta - (1/\zeta)} \sqrt{\frac{r'}{r}} J_\zeta(kr_<) H_\zeta^{(1)}(kr_>) \frac{d}{d\theta} P_{-\zeta}^{(2)}(\theta) \frac{d}{d\theta'} P_{-\zeta}^{(2)}(\theta') \frac{D_{-\zeta}^{(1)}(\theta_0)}{\frac{\partial}{\partial \zeta} D_{-\zeta}^{(2)}(\theta_0)} \\ & + \frac{i \omega \epsilon \mathcal{M} \sin \theta'}{2} \sqrt{\frac{r'}{r}} \left\{ \int_0^{\infty + i\epsilon} \frac{\mu}{\mu^2 - \frac{1}{4}} J_\mu(kr_<) H_\mu^{(1)}(kr_>) \frac{\partial^2}{\partial \theta \partial \theta'} \bar{G}_\theta(\theta, \theta'; \mu) d\mu \right. \\ & \left. - \int_0^{\infty - i\epsilon} \frac{\mu}{\mu^2 - \frac{1}{4}} J_\mu(kr_<) H_\mu^{(1)}(kr_>) \frac{\partial^2}{\partial \theta \partial \theta'} \bar{G}_\theta(\theta, \theta'; -\mu) d\mu \right\}, \quad (40) \end{aligned}$$

$$D_\mu^{(\alpha)}(\theta_0) = \frac{d}{d\theta_0} P_\mu^{(\alpha)}(\theta_0) + c(\mu^2 - \frac{1}{4}) P_\mu^{(\alpha)}(\theta_0), \quad \alpha = 1, 2, \quad (40a)$$

the simple poles, $\mu_p = -n - \frac{1}{2}$, $n = 0, 1, 2, \dots$, of the gamma function $\Gamma(\mu + \frac{1}{2})$ in (36b), one notes that the only singularities of \bar{G}_θ in the right half of the μ plane for $\text{Im } \mu \neq 0$ are the possible poles of $\hat{\Gamma}$. To facilitate the investigation of these poles, we assume that $|\mu|$ is so large that the asymptotic formula, [cf. (26b)]

$$P_\mu^{(1),(2)}(\theta) \sim \sqrt{\frac{2}{\pi \mu \sin \theta}} e^{\pm i(\mu \theta - \pi/4)},$$

$$|\mu| \sin \theta \gg 1, \quad |\arg \mu| < \pi, \quad (38)$$

can be employed. Under these conditions,

$$\hat{\Gamma} \cong \frac{\mu + i\eta}{\mu - i\eta} i e^{i2\mu\theta_0}, \quad \text{Im } \mu > 0, \quad (39a)$$

$$\eta = \frac{1}{c}.$$

$$\hat{\Gamma} \cong \frac{\mu - i\eta}{\mu + i\eta} i e^{-i2\mu\theta_0}, \quad \text{Im } \mu < 0, \quad (39b)$$

Since $\text{Im } c \leq 0$, i.e., $\text{Im } \eta \geq 0$, one notes that the only possible pole in the right-half μ plane exists in the fourth quadrant at $\mu_p = -i\eta$, and only when $\text{Re } \eta > 0$. (Note: the case $\text{Im } \eta = 0$ is obtained by first considering $\text{Im } \eta \neq 0$ and then letting $\text{Im } \eta \rightarrow 0$.) This pole corresponds to the surface wave, and from physical considerations about the existence of a surface wave on a reactive surface, the occurrence of a pole in the fourth quadrant only, and only when $\text{Re } c > 0$, must hold even when the restriction $|c|$ small, implied in (39), is removed.

We can now obtain a Green's function representation analogous to that in (21) by deforming the contour P in the integral representation for \bar{S} as in (27a) about the real axis and extracting the residue at the pole. The expression for the magnetic field \bar{H}_ϕ is then given from (27) by

$$D_{-\zeta}^{(2)}(\theta_0) = 0. \quad (40b)$$

The first term in (40) yields the surface wave contribution and exists only if $\text{Re } c > 0$. The terms involving the integrals represent the contribution from the continuous spectrum; the contour for the first integral runs slightly above the real μ axis, while that for the second integral runs slightly below. If $|c|$ is small, $\zeta \approx -i/c$, and the surface wave term reduces from (38) to that in (35).

It is also of interest to note that the dominant contribution to the integrals in (40) for large r and r' arises from the large- μ range of the integrand. (This dominant contribution represents the geometric-optical field.) If one approximates $\bar{G}_\theta(\theta, \theta'; \pm\mu)$ by its first-order large- μ asymptotic form from (38), one finds that the integrals in (40) become identical with that in (21). In fact, the approximate relation,

$$\bar{H}_\phi|_{\text{cone}} \cong \sqrt{\frac{r' \sin \theta'}{r \sin \theta}} \bar{H}_s|_{\text{wedge}}, \quad (41)$$

is found to apply if the angular functions in (40) are approximated by their first order asymptotic value. Eq. (41) states the quasi-optic result that the azimuthal magnetic field excited by a large magnetic-current ring source in the presence of the cone configuration is approximately equal to the longitudinal magnetic field excited by a magnetic current line source (of the same strength per unit length) in the presence of the wedge configuration, multiplied by the divergence factor $(r' \sin \theta' / r \sin \theta)^{1/2}$, provided that the variables $r, (\theta_0 - \theta), (\theta_0 - \theta')$ for the cone are identified with the variables ρ, ϕ, ϕ' for the wedge. (Diffraction effects are neglected in this equivalence.) An investigation of a

general relation between azimuthally symmetric field problems in spherical coordinates and longitudinally independent problems in cylindrical coordinates has been given elsewhere.²⁰

APPENDIX I

The Angular Spectrum

A) *Wedge Problem*: The angular eigenfunctions $\Phi_\xi(\phi)$ satisfy the differential equation,

$$\left(\frac{d^2}{d\phi^2} + \lambda_\xi\right)\Phi_\xi(\phi) = 0, \quad 0 \leq \phi \leq \alpha, \quad \lambda_\xi = \xi^2, \quad (42)$$

subject to the boundary conditions,

$$\Phi_\xi = \mp c_{0,\alpha} \frac{d\Phi_\xi}{d\phi} \text{ at } \Phi = 0, \alpha, \quad \text{Im } c_{0,\alpha} \leq 0. \quad (42a)$$

c_0 and c_α are constants. Upon multiplying Eq. (42) by Φ_ξ^* (* denotes the complex conjugate), integrating over ϕ between $\phi=0$ and $\phi=\alpha$, and employing integration by parts and the boundary conditions (42a), one obtains the following expression for the eigenvalue λ_ξ :

$$\lambda_\xi = \frac{A - \frac{1}{c_0} B(0) - \frac{1}{c_\alpha} B(\alpha)}{C}, \quad (43)$$

where

$$A = \int_0^\alpha \left| \frac{d\Phi_\xi}{d\phi} \right|^2 d\phi, \quad B(\phi) = |\Phi_\xi(\phi)|^2, \quad (43a)$$

$$C = \int_0^\alpha |B(\phi)|^2 d\phi. \quad (43a)$$

Since A , B , and C are positive quantities (or zero) and $\text{Im}(1/c_{0,\alpha}) \geq 0$, it follows that $\text{Im } \lambda_\xi \leq 0$, and $\text{Re } \lambda_\xi$ may become negative only if either $\text{Re } c_0 > 0$, or $\text{Re } c_\alpha > 0$, or both. For real $c_{0,\alpha}$, λ_μ is real. The eigenfunctions $\Phi_\xi(\phi)$ satisfy the orthogonality condition,

$$\int_0^\alpha \Phi_\xi(\phi) \Phi_{\xi'}(\phi) d\phi = 0, \quad \lambda_\xi \neq \lambda_{\xi'}. \quad (44)$$

B) *Cone Problem*: The eigenfunctions $\Theta_\xi(\theta) = P_{\xi-1/2}(\cos \theta)$ in (34) satisfy the differential equation,

$$\left(\frac{d}{d\theta} \sin \theta \frac{d}{d\theta} + \lambda_\xi \sin \theta\right) \Theta_\xi(\theta) = 0, \quad (45)$$

$$0 \leq \theta \leq \theta_0, \quad \lambda_\xi = \xi^2 - \frac{1}{4}, \quad (45)$$

subject to the boundary conditions,

$$\Theta_\xi \text{ non-infinite at } \theta = 0,$$

$$\frac{d\Theta_\xi}{d\theta} = -c\lambda_\xi \Theta_\xi \text{ at } \theta = \theta_0, \quad \text{Im } c \leq 0. \quad (45a)$$

Upon proceeding as above one finds that

$$\frac{1}{\lambda_\xi} = \frac{\hat{A} - c \sin \theta_0 \hat{B}(\theta_0)}{\hat{C}}, \quad (46)$$

$$\hat{A} = \int_0^{\theta_0} \sin \theta \hat{B}(\theta) d\theta, \quad \hat{B}(\theta) = |\Theta_\xi(\theta)|^2, \quad (46a)$$

$$\hat{C} = \int_0^{\theta_0} \sin \theta \left| \frac{d\Theta_\xi}{d\theta} \right|^2 d\theta. \quad (46a)$$

$\sin \theta \geq 0$ in the range $0 \leq \theta \leq \theta_0 \leq \pi$ so that \hat{A} , \hat{B} and \hat{C} are positive (or zero). Since $\text{Im } c \leq 0$, one notes that $\text{Im } \lambda_\xi \leq 0$. Moreover, $\text{Re } \lambda_\xi$ can be negative only if $\text{Re } c > 0$.

However, since the boundary condition at θ_0 in (45a) depends on λ_ξ , the functions Θ_ξ do not satisfy the usual orthogonality relation. In fact, one readily shows that¹⁰

$$\int_0^{\theta_0} \sin \theta \Theta_\xi \Theta_{\xi'} d\theta - \sin \theta_0 c \Theta_\xi(\theta_0) \Theta_{\xi'}(\theta_0) = 0. \quad (47)$$

Eq. (47) can be written as

$$\int_0^{\theta_0} \sin \theta \Theta_\xi(\theta) \varphi_{\xi'}(\theta) d\theta = 0, \quad (48)$$

where

$$\varphi_{\xi'}(\theta) = \Theta_{\xi'}(\theta) [1 - c\delta(\theta_0 - \theta)], \quad (48a)$$

and we have defined

$$\int_0^{\theta_0} f(\theta) \delta(\theta_0 - \theta) d\theta$$

$$= \lim_{\alpha \rightarrow \theta_0} \int_0^\alpha f(\theta) \delta(\alpha - \theta) d\theta = f(\theta_0), \quad 0 < \alpha < \theta_0. \quad (48b)$$

Thus, one may interpret (48) as an orthogonality relation between the functions Θ_ξ and φ_ξ .

APPENDIX II

Conditions on the Edge (Tip) Behavior of the Fields

A) *Wedge*: The edge condition, requiring finite-energy content in any finite volume τ surrounding the edge of the wedge,²¹ implies that

$$\int_\tau [|\mathbf{E}|^2 + |\mathbf{H}|^2] d\tau = \text{finite}. \quad (49)$$

Since $d\tau = \rho d\rho d\phi dz$ in cylindrical coordinates, any component of \mathbf{E} or \mathbf{H} can behave at worst like $\rho^{-1+\delta}$, $\delta > 0$, as $\rho \rightarrow 0$. The condition $\text{Im } c_{0,\alpha} < 0$ (i.e., a lossy surface

²⁰ L. B. Felsen, "Diffraction by Objects in a Certain Variable Plasma Medium," Microwave Res. Inst., Polytechnic Inst. of Brooklyn, Brooklyn, N. Y., Electrophysics Group Memo., no. 55; July 20, 1959.

²¹ J. Meixner, "Die Kantenbedingung in der Theorie der Beugung Elektromagnetischer Wellen an Vollkommen Leitenden Ebenen Schirmen," *Ann. Phys.*, vol. 6, pp. 2-9; 1949.

impedance) implies $\text{Re } \xi > 0$, $\text{Im } \eta > 0$ in (18), and since $J_x(k\rho) \sim \rho^x$ as $\rho \rightarrow 0$, one notes from (3) and (4) that the edge condition is met. For the lossless case $\text{Im } c_{0,\alpha} = 0$, one has both ξ and η positive; for the surface wave, $|J_{-\eta}(k\rho)|$ is constant as $\rho \rightarrow 0$, corresponding to $\delta = 0$ in the condition above. Thus, in order to satisfy the edge condition for the surface wave, the condition $\text{Im } c_{0,\alpha} = 0^-$ must be imposed, where 0^- is an arbitrarily small negative quantity. This restriction is to be understood for the various investigations in the text where it has been supposed that $c_{0,\alpha}$ is purely real. The results are not thereby noticeably affected.

B) Cone: The "tip condition" for the cone is again given by (49) except that τ now represents a finite volume surrounding the cone tip ($r = 0$). Since $d\tau = r^2 \sin \theta dr d\theta d\phi$ in spherical coordinates, no component of \mathbf{E} or \mathbf{H} can grow more rapidly than $r^{-(3/2)+\delta}$, $\delta > 0$, as $r \rightarrow 0$. An examination of (22), (24), (27), (34) shows that the tip condition is met if $\text{Im } c = 0^-$. When one assumes that $\text{Im } c = 0$, the tip condition is violated for the surface wave (but not for the remaining modes). The restriction $\text{Im } c = 0^-$ is therefore implied in any considerations carried out in the text on the assumption that c is purely real. The results are not thereby noticeably affected.

The Finite Range Wiener-Hopf Integral Equation and a Boundary Value Problem in a Waveguide

RAJ MITTRA†

Summary—In this paper the boundary value problem of a finite bifurcation in a rectangular waveguide is formulated in terms of a finite range Wiener-Hopf integral equation, and the solution of the integral equation is presented.

As a first step, an infinite set of simultaneous equations is obtained from the integral equation. The solution of the above set of equations is then obtained by analytic means.

It is indicated that the method developed can be applied to certain other problems of mathematical physics and a number of examples of such problems are included.

I. INTRODUCTION

IN THIS paper, we shall formulate the problem of a finite bifurcation in a rectangular waveguide, the geometry of which is shown in Fig. 1. The formulation will be in terms of a finite range Wiener-Hopf integral equation.

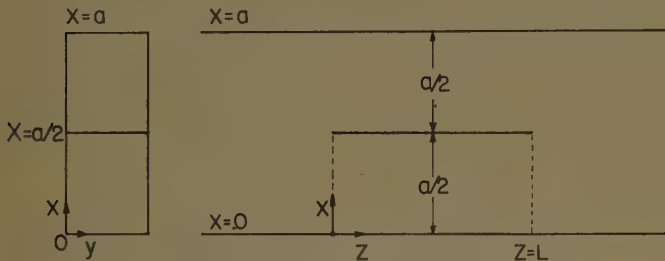


Fig. 1—The finite bifurcation problem in a waveguide.

We shall reduce this problem to the solution of an infinite set of simultaneous linear algebraic equations. By inverting analytically an infinite matrix corresponding to the case when L is infinitely large, a solution for L finite is developed in the form of an infinite series convergent for all $L > 0$. The present problem can also be attacked by the use of the Wiener-Hopf technique as has been pointed out by Noble¹ and by the author.²

The mathematical problem of solving a finite range integral equation of the type,

$$g(y) = \int_0^D f(x) K(|x - y|) dx \quad (1)$$

where $g(y)$ is prescribed for $0 < y < D$, and where

$$K(|x|) = \sum_{n=1}^{\infty} C_n e^{-k_n |x|} \quad (2)$$

has been discussed in considerable detail in a technical

report.² In this paper we shall be concerned only with the particular kernel associated with the present problem which also is a series of the type given in (2). However, solutions to other problems, as for instance that of the surface wave propagation on a corrugated surface as shown in Fig. 2, can also be obtained by the method outlined here.



Fig. 2—Corrugated surface.

II. FORMULATION OF THE PROBLEM

Consider the simplified problem in which the incident electric field in the waveguide is the dominant mode field and is entirely in the y direction. Let it be assumed that only the dominant mode propagates in the guide.³ With no variation of the fields along y , the only field components present are E_y , H_x and H_z and the entire field can be derived from a scalar field potential $A(x, z) = E_y$. The other field components are given by

$$H_x = -\frac{i\omega\epsilon}{k^2} \frac{\partial A}{\partial z}$$

$$H_z = \frac{i\omega\epsilon}{k^2} \frac{\partial A}{\partial x}$$

which are easily verified.

The problem can be stated as follows. The function $A(x, z)$ is to be found, satisfying

$$\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial z^2} + k^2 A = 0 \quad (3)$$

and the following conditions:

- 1) $A = 0$ at $x = 0, a$ for all z ,
- 2) $A = 0$ at $x = a/2$, $0 < z < L$,
- 3) $A(x, z) = \text{incident wave} + \text{reflected wave}$ for $z \rightarrow -\infty$
 $= \text{transmitted wave}$ for $z \rightarrow +\infty$,
- 4) A and ∇A are continuous everywhere except at the edges where $|\nabla A|$ become infinite as $r^{-1/2}$ (r is the radial distance from the edge).

In order to derive the integral equation define a Green's function G satisfying

$$\frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial z^2} + k^2 G = -\delta(x - x_0)\delta(z - z_0), \quad (4)$$

³ This condition can be relaxed without requiring any modification of the results.

† Dept. of Elec. Engrg., University of Illinois, Urbana, Ill.

¹ B. Noble, "The Wiener-Hopf Technique," Pergamon Press, London, Eng., Ch. 5 and p. 174, ex. 4.11; 1958.

² R. Mittra, "On the Solution of a Class of Wiener-Hopf Integral Equations in Finite and Infinite Ranges," Antenna Lab., University of Illinois, Urbana. Tech. Rep. No. 37; 1959.

and the conditions:

- 1) G is continuous except at $x = x_0, z = z_0$ where it has a logarithmic singularity
- 2) G is outgoing for $|z| \rightarrow \infty$,
- 3) $G = 0$ at $x = 0, a$.

Using standard techniques, such a G is easily derived and the Green's function is given by

$$G = -\frac{i}{a} \sum_{n=1}^{\infty} \frac{1}{\gamma_n} \sin n \frac{\pi x}{a} \sin n \pi \frac{x_0}{a} e^{-i\gamma_n |z-z_0|} \quad (5)$$

where

$$\begin{aligned} \gamma_n &= \sqrt{k^2 - \frac{n^2 \pi^2}{a^2}}, & k &> n\pi/a \\ &= -i \sqrt{\frac{n^2 \pi^2}{a^2} - k^2}, & k &< n\pi/a. \end{aligned}$$

The γ_n 's are quickly recognized as the mode propagation constant in the guide.

From (3) and (4) the following is obtained:

$$\int_S (G \nabla^2 A - A \nabla^2 G) dx dz = A(x_0, z_0) \quad (6)$$

where

$$\nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} \right)$$

and the area S is the entire region concerned, i.e., the region bounded by the contour C shown. The surface integral is transformed in a line integral with the contour shown in Fig. 3 and there results after some manipulation,

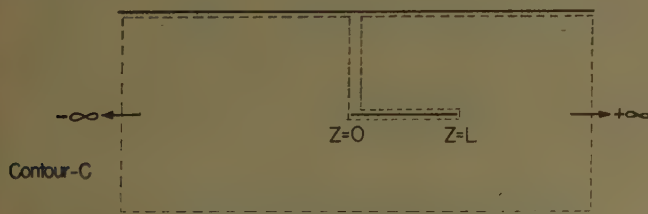


Fig. 3—Contour C and the region of integration.

$$\begin{aligned} \int_S (G \nabla^2 A - A \nabla^2 G) dS &= \int_C \left(G \frac{\partial A}{\partial n} - A \frac{\partial G}{\partial n} \right) dl \\ &= e^{-i\gamma_1 z} \cdot \sin \frac{\pi x}{a} + \int_0^L \{ G \cdot B(x, z) \} \Big|_{x=a/2} dz \end{aligned} \quad (7)$$

where n = outward normal to C and

$$B(x, z) = \frac{\partial A}{\partial x} \Big|_{x=a/2-} - \frac{\partial A}{\partial x} \Big|_{x=a/2+} = \text{discontinuity of } \frac{\partial A}{\partial x} \text{ at } x = a/2.$$

The contribution of the line integral comes only from the plate at $x = a/2$ and from the part of the contour at large negative z . The latter gives rise to the term $e^{-i\gamma_1 z} \cdot \sin \pi x/a$ in (7). From (6) and (7) we obtain,

$$\begin{aligned} A(x_0, z_0) &= e^{-i\gamma_1 z} \cdot \sin \frac{\pi x}{a} + \int_0^L G(x, z, x_0, z_0) \Big|_{x=a/2} \cdot B(z) dz \\ B(z) &= B(x, z) \Big|_{x=a/2}. \end{aligned} \quad (8)$$

Now applying the boundary condition $A = 0$ for $x = a/2, 0 < z < L$, we get the desired integral equation for the current density distribution on the septum which is simply related to $B(z)$. The equation is

$$\begin{aligned} -e^{-i\gamma_1 z_0} &= \int_0^L G(x, z, x_0, z_0) \Big|_{x=a/2} \cdot B(z) dz \\ &= -(i/a) \int_0^L \sum_{r=0}^{\infty} \frac{1}{\gamma_{2r+1}} e^{-i\gamma_{2r+1} |z-z_0|} \cdot B(z) dz \end{aligned} \quad (9)$$

or

$$e^{-i\gamma_1 z_0} = i/a \int_0^L \sum_{r=0}^{\infty} \frac{1}{\gamma_{2r+1}} e^{-i\gamma_{2r+1} |z-z_0|} \cdot B(z) dz.$$

We note that the kernel $K|z-z_0|$ of this integral equation is a series of exponentials and our object is to solve for the function $B(z)$.

Before we go on to discuss the case of finite L we shall solve the equation for $L = \infty$. The reason for doing this is twofold. First of all, in so doing, we shall be able to point out the modification necessary to go from the infinite to the finite L case. Furthermore, we shall see that the finite L solutions are useful in developing the expressions for $B(z)$ for a finite L . The next section will therefore be devoted to the case when the length L is infinite in the positive z direction.

III. SOLUTION FOR THE INFINITE L CASE

In this section we discuss the special case of $L = \infty$. It can be shown without any difficulty that the integral equation in the present case is obtained by putting $L = \infty$ in (9) and although conditions are now different for $z \rightarrow \infty$, this does not modify (9) in any way.

Hence, we consider the equation

$$e^{-i\gamma_1 z_0} = \frac{i}{a} \int_0^{\infty} B(z) \sum_{r=0}^{\infty} \frac{1}{\gamma_{2r+1}} e^{-i\gamma_{2r+1} |z-z_0|} \cdot dz. \quad (10)$$

Assume now that $B(z)$ admits an expansion of the form,

$$B(z) = \sum_{n=1}^{\infty} C_n e^{-i\gamma_{2n} z}. \quad (11)$$

Such an assumption is justified by the easily recognized fact that γ_{2n} 's are the mode functions of the half

guide. Another explanation is that γ_{2n} 's are the zeros of the Fourier transform of $K|z|$, a fact which is stated here without proof but can be verified without too much difficulty. As was explained in the technical report,² a series of exponentials of the zeros of the Fourier transform of the kernel is the proper representation for the unknown function of the integral equation.

Presently, the problem is to find the unknown coefficients C_n . We discuss in the following a method for obtaining a set of simultaneous equations for the coefficients C_n . It should be pointed out that this problem (for $L = \infty$) has also been worked out by Marcuvitz⁴ and Hurd and Gruenberg,⁵ but by using different methods.

Going back to (10) and (11), we notice that the substitution of (11) and (10) would involve the evaluation of an integral I of the type,

$$I = \int_0^{\infty} e^{-i\gamma_{2n}z} \cdot e^{-i\gamma_{2r+1}|z-z_0|} dz \quad (12)$$

for various combinations of n and r .

Upon evaluating the right-hand side of (12), there is obtained

$$I = \frac{e^{-i\gamma_{2r+1}z_0}}{i(\gamma_{2n} - \gamma_{2r+1})} + \frac{2\gamma_{2r+1}e^{-i\gamma_{2n}z_0}}{i(\gamma_{2r+1} - \gamma_{2n}^2)} \quad (13)$$

Hence, we conclude the following. The integral in the right-hand side of (10), when evaluated, yields series of exponentials of the type

$$e^{-i\gamma_{2r+1}z_0} \quad (r = 0, 1, \dots) \text{ and } e^{-i\gamma_{2n}z_0} \quad (n = 1, 2, \dots).$$

Now, in order for (10) to be true we must satisfy the following conditions:

- 1) The coefficient of $e^{-i\gamma_{1n}z_0}$ must agree in both sides, i.e., the coefficient of $e^{-i\gamma_{1n}z_0}$ in the right-hand side must be equal to 1,
- 2) the coefficients of $e^{-i\gamma_{2r+1}z_0}$ must equal zero for all $r \neq 0$,
- 3) The coefficients of $e^{-i\gamma_{2n}z_0}$ must equal zero for all n .

The first two conditions yield a set of equations for C_n . The equations are,

$$\sum_{n=1}^{\infty} \frac{C_n}{\gamma_{2n} - \gamma_{2r+1}} = a\gamma_{1n}\delta_0^r, \quad r = 0, 1, \dots \quad (14)$$

Condition 3, however, yields a set of identities as follows:

$$\sum_{r=0}^{\infty} \frac{1}{\gamma_{2r+1}^2 - \gamma_{2n}^2} = 0, \quad n = 0, 1, \dots \quad (15)$$

Clearly, (15) does not impose any condition on C_n 's. Implicitly, however, the equations put a condition on

⁴ N. Marcuvitz, "Waveguide Handbook," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 10; 1951.

⁵ R. Hurd and H. Gruenberg, "H-plane bifurcation of rectangular waveguides," *Canad. J. Phys.*, vol. 32, pp. 694-701; November, 1954.

the exponentials which were used to express $B(z)$ in (11). It can be verified that the particular choice of the mode functions γ_{2n} 's in the series of exponentials for $B(z)$ instead of any other arbitrary set of constants, indeed does make (15) identically true. Another way of looking at it is the following. γ_{2n} 's are the zeros of the Fourier transform of $K|z|$ and hence (10) is identically true.

We therefore should concentrate on solving (14) for C_n . It is seen that we shall accomplish this by calculating Δ_{1n}/Δ where Δ is the determinant of (14) and Δ_{1n} is its cofactor. In the following section we present a method for calculating the C_n 's.

IV. CALCULATION OF THE UNKNOWN COEFFICIENTS FOR $L = \infty$

To calculate the coefficients C_n 's it is useful to consider a finite size matrix, i.e., one with finite n , and then go to the limit of $n \rightarrow \infty$. The determinant of the system, if a $(p \times p)$ one is considered, is

$$\Delta = \begin{vmatrix} \frac{1}{\gamma_2 - \gamma_1} & \frac{1}{\gamma_4 - \gamma_1} & \dots & \frac{1}{\gamma_{2p} - \gamma_1} \\ \frac{1}{\gamma_2 - \gamma_3} & \frac{1}{\gamma_4 - \gamma_3} & \dots & \frac{1}{\gamma_{2p} - \gamma_3} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\gamma_2 - \gamma_{2p-1}} & \dots & \dots & \frac{1}{\gamma_{2p} - \gamma_{2p-1}} \end{vmatrix} \quad (16)$$

The determinant Δ , due to the form of its elements, is called a "double alternant" in the language of the theory of determinants. This is because each element of the determinant is a reciprocal of the difference $(\gamma_{2n} - \gamma_{2n+1})$ and only one index, viz., n , varies as we go along a row and the other index n changes (keeping n constant) when we go along the columns.

Because of its form, the determinant can be simply expressed in terms of difference products as shown below.

$$\Delta = \frac{\prod_{l=q+1, q=0}^{p-1, l-1} (\gamma_{2l+1} - \gamma_{2q+1}) \cdot \prod_{l=q+1, q=1}^{p-1, l-1} (\gamma_{2q} - \gamma_{2l})}{\prod_{l=1, q=0}^{p, p-1} (\gamma_{2l} - \gamma_{2q+1})} \quad (17)$$

Notice first of all that the expression for Δ in (17) contains all the factors $(\gamma_{2l} - \gamma_{2q+1})$ in the denominator which it must, in view of the nature of the elements of the determinant given by (16). Moreover, if one lets, say, $\gamma_2 = \gamma_4$ in (16) it is seen that $\Delta = 0$. Hence Δ must have combinations of factors of the type $(\gamma_{2q} - \gamma_{2l})$. For similar reasons it also has factors of the type $(\gamma_{2l+1} - \gamma_{2q+1})$. A little algebra of repeated subtraction of rows from rows and columns from columns would show that (17) is indeed true.

The calculation of the cofactors follows along the same lines. We note for instance that Δ_{1t} is just a $(p-1) \times (p-1)$ determinant of the same type as Δ , so we have,

$$\Delta_{1t} = (-1)^{1+t}$$

$$\frac{\prod_{l=q+1, q=1}^{p-1, l-1} (\gamma_{2l+1} - \gamma_{2q+1}) \prod_{l=q+1, q=1}^{p, l-1} (\gamma_{2q} - \gamma_{2l})}{\prod_{l=1, q=1}^{p, p-1} (\gamma_{2l} - \gamma_{2q+1})} \quad (18)$$

where the prime on the continued product implies that the factor corresponding to $l=t$ is to be omitted. Similarly, the double prime implies that the factor corresponding to $q=t$ is also omitted.

Essentially then, we get Δ_{1t} from Δ by cancelling out the factors containing γ_1 and γ_{2t} and using a positive or negative sign depending on whether t is odd or even, respectively. Since our primary interest is to form Δ_{1t}/Δ , we calculate it using (17) and (18). By taking the appropriate ratio and simplifying, we obtain,

$$\frac{\Delta_{1t}}{\Delta} = (-1)^{1+t} \frac{\prod_{q=0}^{p-1} (\gamma_{2t} - \gamma_{2q+1}) \prod_{l=1}^p (\gamma_{2l} - \gamma_1)}{\prod_{l=1}^{p-1} (\gamma_{2l+1} - \gamma_1) \prod_{l=t+1}^p (\gamma_{2t} - \gamma_{2l}) \prod_{l=1}^{t-1} (\gamma_{2l} - \gamma_{2t})} \quad (19)$$

Noting that the last two factors in the denominator can be combined and rewritten as

$$\prod_{l=t+1}^p (\gamma_{2t} - \gamma_{2l}) \prod_{l=1}^{t-1} (\gamma_{2l} - \gamma_{2t}) = (-1)^{t-1} \prod_{l=1}^n (\gamma_{2t} - \gamma_{2l}), \quad (20)$$

the ratio Δ_{1t}/Δ is simply expressed as,

$$\frac{\Delta_{1t}}{\Delta} = \frac{\prod_{q=0}^{p-1} (\gamma_{2t} - \gamma_{2q+1}) \prod_{l=1}^p (\gamma_{2l} - \gamma_1)}{\prod_{l=1}^{p-1} (\gamma_{2l+1} - \gamma_1) \prod_{l=1}^p (\gamma_{2t} - \gamma_{2l})} \quad (21)$$

Eq. (21) gives an expression for Δ_{1t}/Δ in terms of the mode constants of the guide. Once Δ_{1t}/Δ is known in the explicit form of (21), then (14) can be considered as formally solved. We have yet, of course, to let the index p go to infinity. Before we do that, however, let us modify slightly the expression in the right-hand side of (21). Rewrite (21) as,

$$\frac{\Delta_{1t}}{\Delta} = \frac{\prod_{q=0}^{p-1} (\gamma_{2q+1} - \gamma_{2t}) \prod_{l=1}^p (\gamma_{2l} - \gamma_1)}{\prod_{l=1}^{p-1} (\gamma_{2l+1} - \gamma_1) \prod_{l=1}^p (\gamma_{2l} - \gamma_{2t})} \quad (22)$$

and again as,

$$\frac{\Delta_{1t}}{\Delta} = \pi \left[\frac{\prod_{q=0}^{p-1} (\gamma_{2q+1} - \gamma_{2t}) \frac{a}{(2q+1)\pi}}{\prod_{q=1}^p (\gamma_{2q} - \gamma_{2t}) \frac{a}{2q\pi}} \right] \left[\frac{\prod_{q=1}^p (\gamma_{2q} - \gamma_1) \frac{a}{2q\pi}}{\prod_{q=1}^{p-1} (\gamma_{2q+1} - \gamma_1) \frac{a}{(2q+1)\pi}} \right] \quad (23)$$

The reason for introducing the multiplying factors of the type $a/2q\pi$ and $a/(2q+1)\pi$ is the following. When $p \rightarrow \infty$ the ratio of products inside the curly brackets can be asymptotically expressed as a ratio of gamma functions multiplied by a finite number of terms. This makes the numerical calculation easier and, moreover, assures the convergence of the expression in (23). This point is discussed in more detail in the Appendix.

Using (23) in (14) we obtain the expressions for C_n as,

$$C_n = \gamma_1 \pi \cdot \lim_{p \rightarrow \infty} \left[\frac{\prod_{q=0}^{p-1} (\gamma_{2q+1} - \gamma_{2n}) \frac{a}{(2q+1)\pi}}{\prod_{q=1}^p (\gamma_{2q} - \gamma_{2n}) \frac{a}{2q\pi}} \right] \left[\frac{\prod_{q=1}^p (\gamma_{2q} - \gamma_1) \frac{a}{2q\pi}}{\prod_{q=1}^{p-1} (\gamma_{2q+1} - \gamma_1) \frac{a}{(2q+1)\pi}} \right] \quad (24)$$

We have in (24), in effect, the solution of the integral equation for the case of $L = \infty$. In the following we proceed to show that the idea developed in this section can be extended to the finite L case.

V. SOLUTION FOR FINITE L

Let us go back to Section II and to (9) where we have the integral equation,

$$e^{-i\gamma_1 z_0} = \frac{i}{a} \int_0^L \sum_{r=0}^{\infty} \frac{1}{\gamma_{2r+1}} e^{-i\gamma_{2r+1}|z-z_0|} B(z) dz, \quad 0 < z_0 < L,$$

and the problem is to determine $B(z)$. For L finite we assume the following expansion

$$B(z) = \sum_{n=1}^{\infty} D_n e^{-i\gamma_{2n} z} + \sum_{n=1}^{\infty} F_n e^{i\gamma_{2n}(z-L)}. \quad (25)$$

Physically the choice is dictated by the fact that in this case we have to allow for standing waves in the region of $0 < z < L$, and hence include the terms $e^{-i\gamma_{2n}(L-z)}$ because of the extra discontinuity introduced at $z=L$. Mathematically the reason is that both $e^{-i\gamma_{2n}z}$ and $e^{i\gamma_{2n}z}$ are now acceptable because the upper limit L is finite, whereas when L is infinite the integrals involving $e^{i\gamma_{2n}z}$ do not converge.

When (25) is substituted in the integral equation, we get in this case the following four different types of terms after evaluating the integral in that equation.

$$e^{-i\gamma_{2n}z_0}, \quad e^{i\gamma_{2n}(z_0-L)}, \quad e^{-i\gamma_{2r+1}z_0} \quad \text{and} \quad e^{i\gamma_{2r+1}z_0}.$$

Following the same argument as was used for the case of infinite L , one sets the conditions:

- 1) the coefficients of $e^{-i\gamma_{2n}z_0}$ and $e^{i\gamma_{2n}(z_0-L)}$ should equal zero for all n , and
- 2) the coefficients of $e^{i\gamma_{2r+1}z_0}$ should equal zero for all r and of $e^{-i\gamma_{2r+1}z_0}$ for all r except $r=0$.

$$\Delta_L = \begin{vmatrix} \left(\frac{1}{\gamma_2 - \gamma_1} - \frac{e^{-i\gamma_2 L}}{-\gamma_2 - \gamma_1} \right), & \left(\frac{1}{\gamma_4 - \gamma_1} - \frac{e^{-i\gamma_4 L}}{-\gamma_4 - \gamma_1} \right), & \cdots & \cdots \\ \left(\frac{1}{\gamma_2 - \gamma_3} - \frac{e^{-i\gamma_2 L}}{-\gamma_2 - \gamma_3} \right), & \left(\frac{1}{\gamma_4 - \gamma_3} - \frac{e^{-i\gamma_4 L}}{-\gamma_4 - \gamma_3} \right), & \cdots & \cdots \\ \cdots & \cdots & \ddots & \cdots \\ \cdots & \cdots & \cdots & \ddots \end{vmatrix}. \quad (30)$$

Since the left hand side of the integral equation contains only $e^{-i\gamma_{2n}z_0}$, it is seen with little difficulty that the above conditions must be satisfied in order for the two sides to agree for all z_0 in the range concerned.

As in the infinite L case, it is found that the coefficient of $e^{-i\gamma_{2n}z_0}$, when equated to zero, yields an identity. This identity turns out to be the same as in the $L = \infty$ case. Furthermore, the coefficient of $e^{i\gamma_{2n}(z_0-L)}$ equated to zero yields the same identity. Hence condition 1) is automatically satisfied. Equating the coefficients of $e^{-i\gamma_{2r+1}z_0}$ and $e^{i\gamma_{2r+1}z_0}$ yields the following doubly infinite set of equations for the coefficients D_n and F_n :

$$\sum_{n=1}^{\infty} \left(\frac{D_n}{\gamma_{2n} - \gamma_{2r+1}} - \frac{F_n e^{-i\gamma_{2n}L}}{\gamma_{2n} + \gamma_{2r+1}} \right) = a\gamma_1 \delta_0^r, \quad (26)$$

$$\sum_{n=1}^{\infty} \left(\frac{D_n e^{-i\gamma_{2n}L}}{\gamma_{2n} + \gamma_{2r+1}} - \frac{F_n}{\gamma_{2n} - \gamma_{2r+1}} \right) = 0. \quad (27)$$

There doesn't seem to be a way of expressing the determinant of (26) and (27) in a simple form. However, a couple of steps modify things to a more suitable form.

Adding and subtracting (27) and (28) we get,

$$\sum_{n=1}^{\infty} (D_n - F_n) \left[\frac{1}{\gamma_{2n} - \gamma_{2r+1}} - \frac{e^{-i\gamma_{2n}L}}{-\gamma_{2n} - \gamma_{2r+1}} \right] = a\gamma_1 \delta_0^r \quad (28)$$

$$\sum_{n=1}^{\infty} (D_n + F_n) \left[\frac{1}{\gamma_{2n} - \gamma_{2r+1}} + \frac{e^{-i\gamma_{2n}L}}{-\gamma_{2n} - \gamma_{2r+1}} \right] = a\gamma_1 \delta_0^r. \quad (29)$$

It is seen that (28) and (29) are independent sets of equations, and the determinants associated with them are quite similar. We also notice that $L = \infty$ would make the determinants of the sets the same and hence would imply that $D_n - F_n = D_n + F_n$ or simply $F_n = 0$, a result which we should of course expect. We now proceed to develop the expressions for the determinants of (28) and (29) and their cofactors.

Written explicitly, the determinant of (28), say Δ_L is,

It is seen that (30) is a determinant, each column of which is a sum of two terms. One can therefore use the well-known rule for expanding such a determinant. Consider for instance the matrix $[M]$ which can be written as a sum of two matrices $[M_1]$ and $[M_2]$. Let

$$[M] = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix} = [M_1] + [M_2] \quad (31)$$

where

$$[M_1] = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad [M_2] = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

Then the determinant Δ of M can be expressed as

$$\Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & b_{12} \\ a_{21} & b_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & a_{12} \\ b_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix}. \quad (32)$$

It is seen that the determinant Δ can be expressed as a sum of, say, subdeterminants which are formed by picking all possible combinations of the columns of the matrices $[M_1]$ and $[M_2]$ but never changing the respective positions of the columns.

The rule is easily extended to an arbitrary size determinant. Going back to (30), it is noticed that the matrix associated with it, say $[N]$ can also be expressed as a sum of two matrices $[N_1]$ and $[N_2]$ as follows:

$$[N] = [N_1] + [N_2]$$

$$= \begin{bmatrix} \frac{1}{\gamma_2 - \gamma_1}, & \frac{1}{\gamma_4 - \gamma_1}, & \dots & \dots & \dots \\ \frac{1}{\gamma_2 - \gamma_3}, & \frac{1}{\gamma_4 - \gamma_3}, & \dots & \dots & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix} + \begin{bmatrix} \frac{e^{-i\gamma_2 L}}{-\gamma_2 - \gamma_1}, & \frac{e^{-i\gamma_4 L}}{-\gamma_4 - \gamma_1}, & \dots & \dots & \dots \\ \frac{e^{-i\gamma_2 L}}{-\gamma_2 - \gamma_3}, & \frac{e^{-i\gamma_4 L}}{-\gamma_4 - \gamma_3}, & \dots & \dots & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}. \quad (33)$$

Notice now that the first matrix N_1 is the same as the one for $L = \infty$, and we have already calculated its determinant Δ . Notice also that if the first column $[N_1]$ is replaced by the first column of $[N_2]$, the new determinant can be expressed as $-e^{-i\gamma_2 L} \Delta(-\gamma_2)$, where $\Delta(-\gamma_2)$ is the original determinant Δ of $[N_1]$ with γ_2 replaced by $-\gamma_2$. Thus we have the following possible expansion for the determinant Δ_L :

$$\Delta_L = \Delta \left\{ 1 - \sum_{q=1}^{\infty} e^{-i\gamma_{2q} L} \Delta(-\gamma_{2q}) / \Delta \right. \\ + \sum_{q=1, r=q+1}^{\infty, \infty} e^{-i(\gamma_{2q} + \gamma_{2r}) L} \Delta(-\gamma_{2q}, -\gamma_{2r}) / \Delta \\ - \sum_{q=1, r=q+1, s=r+1}^{\infty, \infty, \infty} e^{-i(\gamma_{2q} + \gamma_{2r} + \gamma_{2s}) L} \Delta(-\gamma_{2q}, -\gamma_{2r}, -\gamma_{2s}) / \Delta \\ \left. + \dots - \dots \right\}. \quad (34)$$

Since the higher order mode coefficients γ_{2q} 's, etc. are negative imaginary numbers and $\gamma_{2q} \rightarrow -2q(\pi i/a)$ as q becomes large, it is seen that the higher order terms tend to zero in an exponential manner. It can be verified from the expressions of the ratios of the determinants in (34) that they only have an algebraic rather than exponential behavior.

The cofactors of the determinant, say Δ_{1nL} , can be expressed in a similar manner. Take for instance Δ_{1tL} . It has the expansion

$$\Delta_{1tL} = \Delta_{1t} \left\{ 1 - \sum_{q=1}^{\infty} e^{-i\gamma_{2q} L} \Delta_{1t}(-\gamma_{2q}) / \Delta_{1t} \right. \\ + \sum_{q=1, r=q+1}^{\infty, \infty} e^{-i(\gamma_{2q} + \gamma_{2r}) L} \Delta_{1t}(-\gamma_{2q}, -\gamma_{2r}) / \Delta_{1t} \\ \left. - \dots + \dots \right\}. \quad (35)$$

The prime on the summation implies that the term corresponding to $q=t$, $r=t$, etc. are to be excluded from the summation. Δ_{1t} is the corresponding cofactor of the determinant Δ , an expression for which was given earlier in (18). The exponential decay of the higher order terms is also present in the expansion (35) of Δ_{1tL} . Using (34) and (35) one has,

$$D_t - F_t = a\gamma_1 \frac{\Delta_{1t}}{\Delta} \frac{U}{V} \quad (36)$$

where U and V are the expressions inside the curly brackets in (35) and (34), respectively.

From the form of the determinant of (29), it is obvious that $(D_t + F_t)$ can be expressed in a similar manner as,

$$(D_t + F_t) = a\gamma_1 \frac{\Delta_{1t}}{\Delta} \frac{X}{Y} \quad (37)$$

where,

$$X = \left\{ 1 + \sum_{q=1}^{\infty} e^{-i\gamma_{2q} L} \Delta_{1t}(-\gamma_{2q}) / \Delta_{1t} \right. \\ + \sum_{q=1, r=q+1}^{\infty, \infty} e^{-i(\gamma_{2q} + \gamma_{2r}) L} \Delta_{1t}(-\gamma_{2q}, -\gamma_{2r}) / \Delta_{1t} \\ \left. + \dots + \dots \right\} \quad (38)$$

and,

$$Y = \left\{ 1 + \sum_{q=1}^{\infty} e^{-i\gamma_{2q} L} \Delta(-\gamma_{2q}) / \Delta \right. \\ + \sum_{q=1, r=q+1}^{\infty, \infty} e^{-i(\gamma_{2q} + \gamma_{2r}) L} \Delta(-\gamma_{2q}, -\gamma_{2r}) / \Delta \\ \left. + \dots + \dots \right\}. \quad (39)$$

Hence, from (36) and (37)

$$D_t = \frac{a}{2} \gamma_1 \frac{\Delta_{1t}}{\Delta} \left[\frac{X}{Y} + \frac{U}{V} \right] \\ F_t = \frac{a}{2} \gamma_1 \frac{\Delta_{1t}}{\Delta} \left[\frac{X}{Y} - \frac{U}{V} \right]. \quad (40)$$

The formal solution of the finite range integral equation is thus complete.

VI. THE COMPLEMENTARY INTEGRAL EQUATION

In this section we shall formulate the complementary integral equation in terms of the fields outside of the bifurcation region. The advantage of using this equation for the determination of the expressions of the fields outside the bifurcation region will be pointed out. In addition to that, the method developed in Section V will be shown to be applicable for solving this equation also. It will therefore be of interest to discuss the complementary integral equation. Consider the geometry of the problem again, as shown in Fig. 4. From the sym-

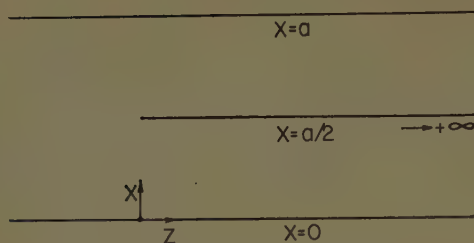


Fig. 4—The semi-infinite bifurcation.

metry of the structure and for the dominant mode incidence it is seen that the potential A satisfies the boundary condition,

$$\frac{\partial A}{\partial x} = 0, \quad z < 0 \quad \text{and} \quad z > L$$

in addition to the ones mentioned previously, *viz.*,

$$A = 0, \quad x = 0, a \text{ for all } z \quad \text{and} \quad A = 0, \quad x = a/2, \quad 0 < z < L.$$

Now suppose we consider the region bounded by the contour C_1 shown in Fig. 5. Choosing a Green's function G which satisfies,

$$\nabla^2 G + k^2 G = -\delta(x - x_0)\delta(z - z_0) \quad (41)$$

$$G = 0, \quad x = \frac{a}{2}, \quad 0 \text{ for all } z$$

and

$$G \rightarrow 0 \text{ as } |z| \rightarrow \infty,$$

the following equation is easily derived:

$$A(x_0, z_0) = - \int_{-\infty}^0 A(x, z) \Big|_{x=a/2} \frac{\partial G}{\partial x} \Big|_{x=a/2} dz - \int_L^{\infty} A(x, z) \Big|_{x=a/2} \frac{\partial G}{\partial x} \Big|_{x=a/2} dz. \quad (42)$$

The Green's function G satisfying (41) is

$$G = -\frac{i}{a} \sum_{n=1}^{\infty} \frac{1}{\gamma_{2n}} e^{-i\gamma_{2n}|z-z_0|} \cdot \sin \frac{2n\pi x}{a} \sin \frac{2n\pi x_0}{a}. \quad (43)$$

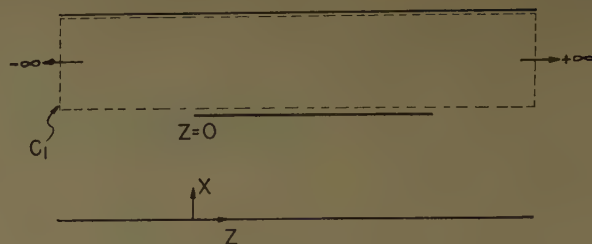
Eq. (42) is the complementary integral equation since the range of integration along z in (42) is complementary to the range of (8). Note also that the unknown now is A , or for that matter the E_y field along z in appropriate ranges rather than the current distribution on the septum.

Applying the conditions

$$\frac{\partial A}{\partial x} \Big|_{x=a/2} = 0$$

for $-\infty < z < 0$ and $L < z < \infty$ we get

$$0 = \int_{-\infty}^0 A \Big|_{x=a/2} \frac{\partial^2 G}{\partial x_0 \partial x} \Big|_{x=a/2} dz + \int_L^{\infty} A \Big|_{x=a/2} \frac{\partial^2 G}{\partial x_0 \partial x} \Big|_{x=a/2} dz - \infty < z_0 < 0 \quad \text{and} \quad L < z_0 < \infty. \quad (44)$$

Fig. 5—Alternate contour C_1 .

Using the expression for $\partial^2 G / \partial x \partial x_0$ in the appropriate ranges in (44), we obtain, writing $P(z) = A(x, z) |_{x=a/2}$,

$$0 = \int_{-\infty}^0 P(z) \sum_{n=1}^{\infty} \left(\frac{2n\pi}{a} \right)^2 \frac{1}{\gamma_{2n}} e^{-i\gamma_{2n}|z-z_0|} dz + \int_L^{\infty} P(z) \sum_{n=1}^{\infty} \left(\frac{2n\pi}{a} \right)^2 \frac{1}{\gamma_{2n}} e^{-i\gamma_{2n}(z-z_0)} dz - \infty < z_0 < 0. \quad (45)$$

Now assume that $P(z) = A(x, z) |_{x=a/2}$ has the form

$$P(z) = e^{-i\gamma_1 z} + \sum_{n=0}^{\infty} A_n e^{i\gamma_{2n+1} z}, \quad -\infty < z < 0 \\ = \sum_{n=0}^{\infty} B_n e^{i\gamma_{2n+1}(z-L)}, \quad z > L \quad (46)$$

which is obtained by assuming that $A(x, z)$ is expanded in terms of the appropriate mode functions in the guide for the ranges of z concerned.

When (46) is substituted in (45) we get integrals I_1 of the type,

$$I_1 = \int_{-\infty}^0 e^{-i\gamma_{2n+1} z} \cdot e^{-i\gamma_{2n}|z-z_0|} dz$$

which are similar to the integrals I appearing in (12) but with the roles of γ_{2r+1} and γ_{2n} interchanged. We also get integrals I_2 , where,

$$I_2 = \int_L^{\infty} e^{-i\gamma_{2n+1}(z-L)} \cdot e^{-i\gamma_{2n}(z-z_0)} dz.$$

When these integrals are evaluated we get the exponentials $e^{i\gamma_{2r+1}z_0}$ $e^{i\gamma_{2n}z_0}$. Using the ideas developed before, we obtain

$$\sum_{p=0}^{\infty} \left(\frac{A_p}{\gamma_{2p+1} - \gamma_{2n}} - \frac{B_p e^{-i\gamma_{2n}L}}{\gamma_{2p+1} - \gamma_{2n}} \right) = \frac{1}{-\gamma_1 - \gamma_{2n}} \\ n = 1, 2, \dots, \quad (47)$$

by equating the coefficient of $e^{-i\gamma_{2n}z_0}$ to zero, which condition is required for the satisfaction of (45). Furthermore, the coefficient of $e^{i\gamma_{2r+1}z_0}$ turns out to be an identity and hence yields no new information or equation for A_p and F_p . Another set of equations is obtained when the range $L < z_0 < \infty$ is used. The set is

$$\sum_{p=0}^{\infty} \left(\frac{A_p e^{-i\gamma_{2n}L}}{\gamma_{2p+1} + \gamma_{2n}} - \frac{B_p}{\gamma_{2p+1} - \gamma_{2n}} \right) = \frac{e^{-i\gamma_{2n}L}}{-\gamma_1 + \gamma_{2n}}, \\ n = 1, 2, \dots \quad (48)$$

From (47) and (48) one may obtain the following equations which prove to be more suitable to work with:

$$\sum_{p=0} (A_p - B_p) \left(\frac{1}{\gamma_{2n} - \gamma_{2p+1}} + \frac{e^{-i\gamma_{2n}L}}{\gamma_{2p+1} - \gamma_{2n}} \right) = \frac{1}{\gamma_1 + \gamma_{2n}} + \frac{e^{-i\gamma_{2n}L}}{\gamma_1 - \gamma_{2n}}, \quad (49)$$

$$\sum_{p=0} (A_p + B_p) \left(\frac{1}{\gamma_{2n} - \gamma_{2p+1}} - \frac{e^{-i\gamma_{2n}L}}{-\gamma_{2p+1} - \gamma_{2n}} \right) = \frac{1}{\gamma_1 + \gamma_{2n}} - \frac{e^{-i\gamma_{2n}L}}{\gamma_1 - \gamma_{2n}},$$

$$n = 1, 2, \dots, \infty. \quad (50)$$

Notice first of all that the matrix of (49) is the transpose of (29) and the matrix of (50) is the transpose of (28). Hence, the method of dealing with them will be expected to be similar to the one developed earlier. However, the right-hand side of the equations here is different from that obtained in the corresponding case in Section V. Here, the right-hand side, treated as a column vector is obtained by replacing γ_1 by $-\gamma_1$ in the first column of the matrix associated with (49), (50).

Hence, we readily obtain the form of the solution

$$A_r - B_r = \frac{\Delta^{(1)}}{\Delta^{(1)}} (\gamma_{2r+1} \rightarrow -\gamma_1) \quad (51)$$

where $\Delta^{(1)}$ is the determinant of (49) and $\Delta^{(1)}(\gamma_{2r+1} \rightarrow -\gamma_1)$ implies the determinant obtained by replacing γ_{2r+1} by $-\gamma_1$.

Similarly,

$$A_r + B_r = \frac{\Delta^{(2)}(\gamma_{2r+1} \rightarrow -\gamma_1)}{\Delta^{(2)}} \quad (52)$$

where $\Delta^{(2)}$ is associated with (50).

If our primary interest lies in obtaining the reflection and the transmission coefficients, A_0 and B_0 , it is seen that we have to evaluate (51) and (52) for $r=0$ only. That A_0 and B_0 are indeed the reflection and transmission coefficients is seen from (46).

To complete the solution we should present the expression for the ratios of Δ 's. The expressions are developed in the manner developed earlier. Consider for instance $\Delta^{(1)}$ which, written explicitly, is

$$\Delta^{(1)} = \begin{vmatrix} \left(\frac{1}{\gamma_2 - \gamma_1} + \frac{e^{-i\gamma_2 L}}{-\gamma_2 - \gamma_1} \right), & \left(\frac{1}{\gamma_2 - \gamma_3} + \frac{e^{-i\gamma_2 L}}{-\gamma_2 - \gamma_3} \right), & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \left(\frac{1}{\gamma_{2m} - \gamma_1} + \frac{e^{-i\gamma_{2m} L}}{-\gamma_{2m} - \gamma_1} \right), & \left(\frac{1}{\gamma_{2m} - \gamma_3} + \frac{e^{-i\gamma_{2m} L}}{-\gamma_{2m} - \gamma_3} \right), & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \end{vmatrix}. \quad (53)$$

The expansion of $\Delta^{(1)}$ obtained by treating each of its rows as a sum of two is,

$$\Delta^{(1)} = \Delta \left\{ 1 + \sum_{m=1}^{\infty} e^{-i\gamma_{2m} L} \frac{\Delta(-\gamma_{2m})}{\Delta} + \sum_{m=1, n=m+1}^{\infty, \infty} e^{-i(\gamma_{2m} + \gamma_{2n}) L} \frac{\Delta(-\gamma_{2m}, -\gamma_{2n})}{\Delta} + \dots + \dots \right\} \quad (54)$$

where,

$$\Delta = \lim_{L \rightarrow \infty} \Delta^{(1)}. \quad (55)$$

The expression for Δ is the same as given in (17), since it is the same determinant as appears in (16).

It is seen that the leading term of the expansion of $\Delta^{(1)}$ is Δ . It will be of interest to derive the expression for the leading term of the reflection coefficient A_0 . Since

$$\lim_{L \rightarrow \infty} \Delta^{(1)} = \Delta^{(2)} = \Delta, \quad \lim_{L \rightarrow \infty} B_r = 0.$$

Hence,

$$\lim_{L \rightarrow \infty} A_0 = \Delta(\gamma_1 \rightarrow -\gamma_1).$$

From (17),

$$\lim_{L \rightarrow \infty} A_0 = \lim_{p \rightarrow \infty} \prod_{l=1}^{p-1} \left\{ \frac{(\gamma_{2l+1} - \gamma_1)}{(\gamma_{2l+1} + \gamma_1)} \right\} \cdot \prod_{l=1}^p \left\{ \frac{(\gamma_{2l} + \gamma_1)}{\gamma_{2l} - \gamma_1} \right\}. \quad (56)$$

The higher order terms are developed without much difficulty.

VII. THE EDGE CONDITION

As was pointed out in Section II, the field potential has to satisfy the edge condition, which is in fact a condition on $|\nabla A|$ in the vicinity of the edge. In this section we shall indicate a way of verifying that the edge condition is indeed satisfied by the solution obtained. An alternative way of stating the edge condition is that

the current density distribution on the septum tends to infinity as $1/z^{1/2}$ as $z \rightarrow 0_+$ and as $1/(L-z)^{1/2}$ as $z \rightarrow L_-$. We will accomplish this if we show that $B(z)$, defined in connection with (7), has the above behavior. Following Hurd⁵ we know that this will follow if in the expansion of $B(z)$, which is

$$B(z) = \sum_{n=1}^{\infty} D_n e^{-i\gamma_{2n}z} + \sum_{n=1}^{\infty} F_n e^{-i\gamma_{2n}(L-z)},$$

we show that,

$$D_n \rightarrow K_1/n^{1/2} \quad \text{as } n \rightarrow \infty, \quad K_1 = \text{constant}$$

$$F_n \rightarrow K_2/n^{1/2} \quad \text{as } n \rightarrow \infty, \quad K_2 = \text{constant}.$$

It is easily seen that the coefficients D_n and F_n determine the behavior of $B(z)$ as $z \rightarrow 0$ and $z \rightarrow L$, respectively.

From (40), remembering that the leading term of the expansion of both X/Y and U/V is 1, we see that the behavior of D_n as $n \rightarrow \infty$ is primarily determined by Δ_{1n}/Δ . As shown in the Appendix, Δ_{1n}/Δ indeed tends to $k_1/n^{1/2}$ as n becomes indefinitely large.

Considering F_n now, we see from (40) that the leading term in F_n is $R e^{-i\gamma_{2n}L} \Delta_{1n}/\Delta$ where R is a constant. Hence, as $z \rightarrow L_-$, $B(z)$ again goes to infinity as $1/(L-z)^{1/2}$. The integrated current at the edge $z=L_-$ course is much smaller compared to the integrated current density at the edge $z=0_+$ because of the multiplying factor $e^{-i\gamma_2L}$ (note that γ_2 is a negative imaginary number).

VIII. NUMERICAL CALCULATION

Only a brief discussion will be given here on the aspect of numerical calculation. We observe first of all that the computation of the mode coefficients in the various regions involves the calculation of infinite series of terms which contain ratios of infinite products. It was pointed out in earlier sections that only a finite number of terms in the infinite series need be taken because of the exponentially decaying nature of the higher order terms in the series. As far as the ratios of infinite products are concerned, they can be expressed asymptotically to ratios of finite products and appropriate gamma functions as shown in the Appendix. Sample numerical computations have been carried out and they have not been found either difficult or time consuming. Although the digital computers are helpful, a hand computer is also quite efficient for obtaining answers even when L is of the order of $\lambda g/2$ for the dominant mode, under the

assumed condition that only the dominant mode is the propagating one. The method of course works better with a larger L .

IX. CONCLUSION

The finite bifurcation problem in a rectangular waveguide has been formulated in terms of a finite range Wiener-Hopf integral equation. The solution of this equation has been obtained by analytic means. The case when the septum is semi-infinite has also been included and a connection between the methods of solution for the finite and the infinite case has been provided, the finite case being an extension of the infinite one. The problem of numerically calculating the quantities of interest, *viz.*, the mode coefficients, has been discussed briefly. The technique presented here is general and is applicable to other finite range Wiener-Hopf integral equations for which the kernel can be expanded as a series of exponentials of the type $e^{-\alpha_n(z-z_0)}$ where z and z_0 are the variable coordinates.

APPENDIX

The purpose of this section is to recast the form of Δ_{1t}/Δ given in (23) into one involving gamma functions which enables one to show the convergence of the products and also helps study the asymptotic behavior of Δ_{1t}/Δ as t becomes large. The technique illustrated here for Δ_{1t}/Δ can also be applied to the ratios of other determinants that appear in the text.

Let us rewrite (23) as

$$\begin{aligned} \Delta_{1t}/\Delta &= (\pi/a) \cdot \lim_{p \rightarrow \infty} \left\{ \frac{\prod_{q=1, q \neq 2t}^{2p} (i\gamma_q - i\gamma_{2t}) \frac{a}{q\pi}}{\left[\prod_{q=1}^p (i\gamma_{2q} - i\gamma_{2t}) \frac{a}{2q\pi} \right]^2} \right\} \\ &\quad \cdot \lim_{p \rightarrow \infty} \left\{ \frac{\left[\prod_{q=1}^p (i\gamma_{2q} - i\gamma_1) \frac{a}{2q\pi} \right]^2}{\prod_{q=2, q \neq 2t}^{2p} (i\gamma_q - i\gamma_1) \frac{a}{q\pi}} \right\} \\ &= (\pi/a) \cdot P_1 \cdot P_2 \end{aligned} \quad (57)$$

where P_1 and P_2 are the ratios of the products in the first and second curly brackets, respectively.

Now P_1 can be modified and put in the form

$$P_1 = \left\{ \frac{\prod_{q=1, q \neq 2t}^{2p} (i\gamma_q - i\gamma_{2t}) \frac{a}{q\pi}}{\prod_{q=1, q \neq 2t}^{2p} \left(1 - \frac{i\gamma_{2t}a}{q\pi}\right)} \right\} \left[\frac{\left\{ \prod_{q=1, q \neq 2t}^{2p} \left(1 - \frac{i\gamma_{2t}a}{q\pi}\right) \right\}}{\left\{ \prod_{q=1}^p \left(1 - \frac{i\gamma_{2t}a}{2q\pi}\right) \right\}^2} \right] \cdot \left\{ \frac{\prod_{q=1}^p \left(1 - \frac{i\gamma_{2t}a}{2q\pi}\right)}{\prod_{q=1}^p (i\gamma_{2q} - i\gamma_{2t}) \frac{a}{2q\pi}} \right\}^2 = R_1 \cdot R_2 \cdot R_3^2 \quad (58)$$

where R_1, R_2, R_3 are the three ratios of products enclosed by the big brackets in that order.

The reason for putting P_1 in the above form will now be explained. Remembering that $i\gamma_q \rightarrow q\pi/a$, for q large compared to k , it is observed that the ratio,

$$\frac{(i\gamma_q - i\gamma_{2t}) \frac{a}{q\pi}}{\left(1 - \frac{\gamma_{2t}a}{q\pi}\right)} \rightarrow 1 \quad \text{for large values of } q. \quad (59)$$

Hence, instead of going to indefinitely large values of p when computing the first curly bracket R_1 in (58), we

can take only a finite number as the upper limit. For a given accuracy requirement, this upper limit, say P , is easily determined. The same is clearly true for the ratio of the products R_3 . What we must do now is evaluate the ratio of the products appearing at the center, or R_2 . To this end consider the following well known expansion of the inverse of the gamma function, $\Gamma(\alpha)$

$$\frac{1}{\alpha\Gamma(\alpha)} = \lim_{m \rightarrow \infty} m^{-\alpha} \prod_{n=1}^m \left(1 + \frac{\alpha}{n}\right). \quad (60)$$

So for large p we can write the limit, say Q_1 of R_2 , as

$$Q_1 = \lim_{p \rightarrow \infty} \left\{ \frac{\prod_{q=1, q \neq 2t}^{2p} \left(1 - \frac{i\gamma_{2t}a}{q}\right)}{\left[\prod_{q=1}^p \left(1 - \frac{i\gamma_{2t}a}{2\pi q}\right)\right]^2} \right\} = \lim_{p \rightarrow \infty} \frac{\prod_{q=1, q \neq 2t}^{2p} \left\{1 - \left(\frac{i\gamma_{2t}a}{\pi q}\right)^2\right\}}{\prod_{q=1}^p \left\{1 - \left(\frac{i\gamma_{2t}a}{2\pi q}\right)^2\right\}^2} \cdot \frac{\prod_{q=1}^p \left\{1 + \frac{i\gamma_{2t}a}{2\pi q}\right\}^2}{\prod_{q=1}^{2p} \left\{1 + \frac{i\gamma_{2t}a}{\pi q}\right\}} \quad (61)$$

or,

$$Q_1 = 2 \cdot \frac{\sin i\gamma_{2t}a}{\sin^2 i\gamma_{2t} \frac{a}{2}} \cdot \frac{\left(1 - \frac{i\gamma_{2t}a}{2\pi t}\right)}{\left(1 + \frac{i\gamma_{2t}a}{2\pi t}\right)} \cdot 2^{i\gamma_{2t}a/\pi} \frac{\Gamma\left(\frac{i\gamma_{2t}a}{\pi}\right)}{\Gamma^2\left(\frac{i\gamma_{2t}a}{2\pi}\right)}. \quad (62)$$

Similarly the second curly bracket P_2 in (57) may be written as

$$P_2 = \lim_{p \rightarrow \infty} \left\{ \frac{\prod_{q=1}^p (i\gamma_{2t} - i\gamma_1) \frac{a}{2q\pi}}{\prod_{q=1}^p \left(1 - \frac{i\gamma_1 a}{2q\pi}\right)} \right\} \cdot \left\{ \frac{\prod_{q=2, q \neq 2t}^{2p} \left(1 - \frac{i\gamma_1 a}{q\pi}\right)}{\prod_{q=2, q \neq 2t}^{2p} \left(i\gamma_q - \frac{i\gamma_1 a}{q\pi}\right) \frac{a}{q\pi}} \right\} \\ \times \left\{ \frac{\sin^2 \frac{i\gamma_1 a}{2}}{\sin i\gamma_1 a} \cdot \frac{\left(1 + \frac{i\gamma_1 a}{2\pi}\right)}{\left(1 - \frac{i\gamma_1 a}{2\pi}\right)} \cdot \frac{1}{2^{(1-i\gamma_1 a/\pi)}} \cdot \frac{\Gamma^2\left(\frac{i\gamma_1 a}{2\pi}\right)}{\Gamma\left(\frac{i\gamma_1 a}{\pi}\right)} \right\}. \quad (63)$$

Combining all this it is possible to express Δ_{1t}/Δ with a high degree of accuracy as,

$$(a/\pi) \cdot (\Delta_{1t}/\Delta) = \left[\frac{J(\gamma_{2t})}{H^2(\gamma_{2t})} \cdot \frac{H^2(\gamma_1)}{J_1(\gamma_1)} \cdot \frac{\Gamma\left(i\gamma_{2t} \frac{a}{\pi}\right)}{\Gamma\left(i\gamma_1 \frac{a}{\pi}\right)} \cdot \frac{\Gamma^2\left(i\gamma_1 \frac{a}{2\pi}\right)}{\Gamma^2\left(i\gamma_{2t} \frac{a}{2\pi}\right)} \cdot \frac{\cot \frac{i\gamma_{2t}a}{2}}{\cot \frac{i\gamma_1 a}{2}} \right] \\ \times \left[\frac{\left(1 - \frac{i\gamma_{2t}a}{2\pi t}\right)}{\left(1 + \frac{i\gamma_{2t}a}{2\pi t}\right)} \cdot \frac{\left(1 + \frac{i\gamma_1 a}{2\pi t}\right)}{\left(1 - \frac{i\gamma_1 a}{2\pi t}\right)} \cdot \frac{2^{i(\gamma_{2t}-\gamma_1)a/\pi}}{\left(1 - \frac{i\gamma_1 a}{\pi}\right)} \right] \quad (64)$$

where

$$J(\alpha) = \left\{ \frac{\prod_{q=1}^P (i\gamma_q - i\alpha) \frac{a}{q\pi}}{\prod_{q=1, q \neq 2t}^P \left(1 - \frac{i\alpha a}{q\pi}\right)} \right\} \quad (65)$$

$J_1 = J$ with the factors $q = 1$ omitted,

and

$$H(\alpha) = \left\{ \frac{\prod_{q=1}^P (i\gamma_{2q} - i\alpha) \frac{a}{2q\pi}}{\prod_{q=1}^P \left(1 - \frac{i\gamma_{1q} a}{2q\pi}\right)} \right\}. \quad (66)$$

Since the products J and H have a finite upper limit P , it is more convenient to use (64) for calculating the ratio Δ_{1t}/Δ , rather than the expression in (57), which involves the upper limit p tending to infinity.

Next we go on to consider the asymptotic behavior of Δ_{1t}/Δ as $t \rightarrow \infty$. Going back to (64), we see that this amounts to studying the behavior of the function, say T , where

$$T = \frac{\Gamma\left(\frac{i\gamma_{2t}a}{\pi}\right) \cot\left(\frac{i\gamma_{2t}a}{2}\right) \left(1 - \frac{i\gamma_{2t}a}{2\pi t}\right)}{\Gamma^2\left(\frac{i\gamma_{2t}a}{2\pi}\right) \left(1 + \frac{i\gamma_{2t}a}{2\pi t}\right)} \cdot 2^{i\gamma_{2t}a/\pi} \cdot \frac{\left(1 + \frac{i\gamma_{1a}}{2\pi t}\right)}{\left(1 - \frac{i\gamma_{1a}}{2\pi t}\right)} \quad (67)$$

as $\gamma_{2t} \rightarrow -i2t\pi/a \rightarrow \infty$. It is noted that the only other factors that involve t are $J(\gamma_{2t})$ and $H(\gamma_{2t})$, and they consist of finite products and their limit for large t are constants which are easily obtained.

Inserting the asymptotic behavior of the gamma functions for large arguments we get, using

$$\left| \cot(i\gamma_{2t}a/2)(t\pi - i\gamma_{2t}a/2) \right| \rightarrow 1 \text{ as } i\gamma_{2t}a/2 \rightarrow t\pi, \\ \lim_{t \rightarrow \infty} |T| \rightarrow S_1/t^{1/2}$$

where S is a constant. Hence there follows that,

$$\lim_{t \rightarrow \infty} \Delta_{1t}/\Delta \rightarrow S_2/t^{1/2},$$

where S_2 is another constant.

Eq. (68) gives the asymptotic behavior of Δ_{1t}/Δ as $t \rightarrow \infty$. It also demonstrates the validity of the statement made in Section VII of the text that Δ_{1t}/Δ has the $1/t^{1/2}$ behavior as $t \rightarrow \infty$.

We shall close the Appendix with one further remark. For numerical calculation, the most convenient expression for the ratio of the products Δ_{1t}/Δ is the one appearing in (64). This is because the products J and H appearing in that expression have a finite upper limit. As pointed out in the beginning of this section, other ratios of products appearing in the text can also be recast into a similar form.

ACKNOWLEDGMENT

The author wishes to acknowledge the support provided by Wright Field Air Development Center through the Research Grant AF33(616)-6079. It is also a pleasure to acknowledge the facilities of the Antenna Laboratory of the University of Illinois who provided for the preparation of the paper.

Fields in the Neighborhood of a Caustic

IRVIN KAY*

Summary—Starting with Picht's solution of the wave equation in the form of an integral of plane waves over a caustic, we derive expressions for the solution which depend entirely on the geometry of the problem. An asymptotic evaluation of Picht's integral for high frequencies gives the desired result which can have a number of different forms, depending on the precise region where the field is being observed. In particular, a change in the asymptotic field occurs in going from the regular ray field to the caustic or from a regular part of the caustic to a cusp.

I. INTRODUCTION

IN this paper we shall consider a particular class of solutions of the wave equation in two dimensions,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + k^2 n^2 u = 0 \quad (n = \text{constant}), \quad (1)$$

in a form which is especially suited for dealing with the case where the wavelength is asymptotically small. It is well known that in the small wavelength limit of diffraction or propagation phenomena the geometrical features dominate the underlying physical situation. Accordingly, the solutions we shall use are completely determined by geometrical considerations.

The geometry of a field can be described optically by a set of rays which may be regarded as the paths of energy flow. Where the rays converge to a focus, or form an envelope (also known as a caustic), it may be assumed that the energy density or the field amplitude is large. Where there is a region containing no rays at all, the energy does not propagate. Given the ray system, the field amplitude in the small wavelength limit can be computed at most points of space by well-known methods. In fact, the optical field amplitude has been calculated even at points on the smooth part of the caustic,¹ although the field becomes infinite in the small wavelength limit on a caustic. It is known that the dependence of the asymptotic field on the wavelength changes suddenly along a ray as it passes through a point on a caustic, and a change must occur again in a different way where the caustic has a cusp. We shall investigate the behavior of the asymptotic field in those singular regions by means of a class of explicitly given fields whose construction was presented by J. Picht.² The Picht fields are a generalization of one used by Debye³ for investigating the case of a ray system which converges to a point focus.

If any field whatever has a geometrical optics limit where the ray system is regular, there will exist a superposition of Picht fields which, together, have the same asymptotic limit. Moreover, the Picht field will have the same asymptotic limit on the caustics of the ray system as the given field, a fact which can be easily seen from an application of Green's theorem.⁴ Thus, the conclusions we shall reach are actually more general than might be expected at first.

What we desire here is a purely geometrical description of the field, a description which in principle would allow us to find the field amplitudes graphically from a graphic reproduction of the corresponding ray system.

II. THE PICTH SOLUTION

To construct the exact solution of (1) given by Picht, we must first determine the ray system of the corresponding geometrical optics problem. This can be done whether we are interested in diffraction by an obstacle in a homogeneous medium, in which case the rays are straight lines, or in propagation through an inhomogeneous medium, in which case the rays are curved lines. In any event, the ray system will generally form one or more envelopes known as caustics. The caustic curves, on the other hand, completely determine the ray system, because at each point on a caustic there is a tangent ray, and every ray is tangent to one of the caustics.

The geometrical optics approximation fails at the caustics because there the true wave amplitude grows indefinitely with the frequency. Our attention will be fixed upon the immediate neighborhood of the caustic. Thus, if the rays of the system are curved, we can replace them by the ray system composed of straight line rays having the same envelopes as the original system; for infinitesimal distances there is no distinction between straight lines and sufficiently smooth, slowly varying curved lines.

Let \mathbf{P} be the position vector to the point at which the field is to be observed; let $\mathbf{R}(s)$ be the radius vector to the caustic C , and s the arc-length parameter along C . Let $\rho(s)$ be the radius of curvature of C , and let $A(s)$ be a quantity determined by the initial distribution of the field amplitude, e.g., at the source. We have, for the unit tangent vector to C ,

$$\frac{d\mathbf{R}}{ds} = T(s) \quad (|T|^2 = 1). \quad (2)$$

* Institute of Mathematical Sciences, New York University, New York, N. Y.

¹ L. Landau and E. Lifshitz, "The Classical Theory of Fields," Addison-Wesley Press, Inc., Cambridge, Mass., pp. 155-161; 1951.

² J. Picht, "Optische Abbildung," Friedr. Vieweg and Sohn Akt.-Ges., Braunschweig, Germany, ch. 8; 1931.

³ P. Debye, "Das Verhalten von Lichtwellen in der Nähe eines Brennpunktes oder einer Brennlinie," *Ann. Physik*, vol. 30, no. 4, p. 755; 1909.

⁴ I. Kay and J. B. Keller, "Asymptotic evaluation of the field at a caustic," *J. Appl. Phys.*, vol. 25, no. 7, p. 876; 1954.

The solution of Picht, corresponding to the caustic C , is then

$$u(P) = \int_C A(s) \left\{ \exp ik[(P - R(s)) \cdot T(s) + s] \right\} \frac{ds}{\rho(s)}, \quad (3)$$

where we have taken the index of refraction n to be one for convenience (see Fig. 1).

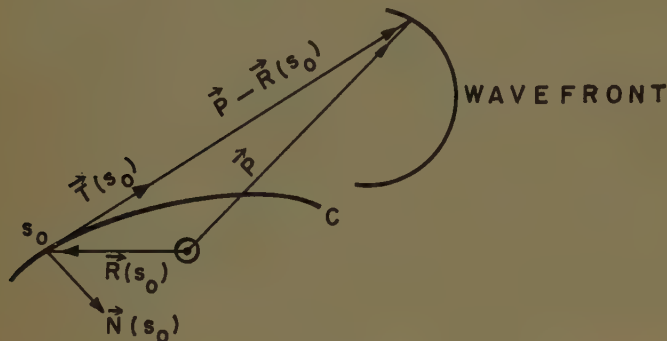


Fig. 1.

The function $\exp ikP \cdot T$, where T is an arbitrary constant vector of unit amplitude, is a plane-wave solution of the wave equation (1). It follows that the integrand in (3) and, thus, that $u(P)$ itself is a solution of (1). Because of the special form of the integrand in (3) $u(P)$ asymptotically approaches the appropriate geometrical optics field at the point P as k becomes large.

III. THE ASYMPTOTIC SOLUTION

The Picht solution (3) is valid everywhere for large values of k insofar as the caustic C determines the field $u(P)$. In this paper we shall investigate in more detail the asymptotic evaluation of (3) for large k so that we may obtain approximate expressions for $u(P)$, which are simple and depend more explicitly upon the ray geometry near P than does (3). In the discussion we shall distinguish between three kinds of regions in which P can be located. The point P is in the "lit" region if a real ray tangent to C at a real point passes through P . If no such ray from C can reach P , we say that P is in the "shadow" region. Finally, P can lie on the caustic C itself.

In the lit region we can evaluate the integral in (3) by stationary phase without running into complications. The derivative with respect to s of the phase of the integrand in (3) is

$$[P - R(s)] \cdot N(s) / \rho(s), \quad (4)$$

where $N(s)$ is the unit normal to the caustic C . Clearly (4) will vanish for $s = s_0$ if the vector $P - R(s_0)$ is parallel to the unit vector $T(s_0)$ tangent to C , i.e., when $P - R(s_0)$ lies along the ray through P . Since P is in the lit region this is possible for real values of the vectors, and the corresponding real value s_0 of the arc-length parameter s is the stationary point. Let us assume that P is on a ray between the source and the stationary point s_0 on the caustic.

The second derivative of the phase or the first derivative of (4) at the stationary point s_0 is

$$-D(s_0) / \rho^2(s_0), \quad (5)$$

where $D(s_0)$ is the magnitude of the vector $P - R(s_0)$.

By expanding the phase in a power series about the stationary point s_0 , and keeping only terms up to the quadratic, we obtain the approximation,

$$u(P) \sim \int_C A(s) \exp \{ ik[D(s_0) + s_0 - D(s_0)(s - s_0)^2 / 2\rho^2(s_0)] \} ds / \rho(s). \quad (6)$$

The important contribution from (6) comes from the immediate neighborhood of s_0 . Thus, we can take the interval of integration from $-\infty$ to $+\infty$ in s without much effect on the result. Now we define

$$\theta - \theta_0 = \int_{s_0}^s ds / \rho(s) \sim (s - s_0) / \rho(s_0)$$

so that $ds / \rho(s) = d\theta$, from which

$$\begin{aligned} u(P) &\sim \{ \exp ik[D(s_0) + s_0] \} \int_{-\infty}^{\infty} A(\theta) \\ &\quad \cdot \exp \{ -ikD(s_0)(\theta - \theta_0)^2 / 2 \} d\theta \\ &\sim A(\theta_0) \{ \exp ik[D(s_0) + s_0] \} \\ &\quad \cdot \int_{-\infty}^{\infty} \exp \{ -ikD(s_0)\theta^2 / 2 \} d\theta. \end{aligned} \quad (7)$$

Finally, we obtain from (7) (see the Appendix)

$$\begin{aligned} u(P) &\sim \sqrt{2\pi / kD(s_0)} A(s_0) \\ &\quad \cdot \exp \{ ik[D(s_0) + s_0] - i\pi/4 \}. \end{aligned} \quad (8)$$

The quantity $D(s_0)$ can be recognized as the radius of curvature of the wavefront through P at the ray which is tangent to the caustic at s_0 , for the caustic is always the locus of the centers of curvature of any wavefront. The equation of the wavefront through P is

$$D(s) + s = D(s_0) + s_0; \quad (9)$$

here $D(s)$ is the distance along the ray at s from the caustic to the wavefront; i.e., the radius of curvature of the wavefront at the wavefront end of the ray through s .

If we set

$$A(s_0) = \hat{A}(\theta_0) \sqrt{k/2\pi} \exp(i\pi/4), \quad (10)$$

then we have

$$u(P) = \hat{A}(\theta_0) D^{-1/2}(s_0) \exp \{ ik[D(s_0) + s_0] \}. \quad (11)$$

The normalized amplitude $\hat{A}(\theta_0)$ will be determined by the angular distribution of energy at the source. We recognize (11) as the geometrical optics solution for the lit region; in particular, note that the amplitude varies inversely as the square root of the radius of curvature of the wavefront.

If we assume that P is in the lit region, not between the source and the caustic, but rather in such a position that the stationary point on the caustic lies on a ray between the source and P , the results at each stage of the calculation are the same except for the sign of the quadratic term of the power series of the phase in (6). The effect is accounted for by multiplying (8) or (11) by $i = \exp(i\pi/2)$. Thus, we have the well-known jump in phase of the geometrical optics solution when a ray passes through the caustic.

When the point P is in the shadow region of the caustic, there is no real value of s on the caustic C for which the vector $P - R(s)$ is parallel to the tangent $T(s)$; *i.e.*, one cannot construct a real tangent to C which passes through P . However, if the radius vector $R(s)$ of C is an analytic function of s , it can be continued into the complex plane analytically. Then, defining

$$\frac{dT(s)}{ds} = N(s)/\rho(s) \quad (12)$$

by analytic continuation for complex values of s , one can generally find a complex value s_0 of s such that the phase is stationary; *i.e.*,

$$[P - R(s_0)] \cdot N(s_0) = 0. \quad (13)$$

The contour C must be deformed until it passes through s_0 . The power series approximation for the phase up to quadratic terms can again be used, and the resulting integral evaluated as before. The approximation will resemble (11) except that, because of the complex phase, an exponential decay factor will be present. Thus, the solution generally decays exponentially in the shadow region.

The quantities involved in this case are geometrical only in an extended sense by analytic continuation. If the caustic C is an arbitrary curve and P is in the neighborhood of C in the shadow region, then we may obtain an approximate evaluation of (3) in terms of real geometrical quantities. This can be done by replacing C by its circle of curvature at the point on C nearest the point P and then evaluating (3) as if the integration contour were the circle of curvature. The evaluation of (3) for the case of a circular caustic will be carried out later.

When P is at a regular point s_0 of the caustic C , s_0 is also a stationary point, because (4) obviously vanishes in this case. However, since D_0 is zero, (5) must also vanish. To obtain the asymptotic evaluation of (3) we must, therefore, use the next nonvanishing term of the power series of the phase in (3). The third derivative of the phase is

$$-\left[\frac{d}{ds}\left(\frac{d\rho}{ds}/(\rho + 1/\rho^3)\right)\right](P - R) \cdot N + 3\left(\frac{d\rho}{ds}/\rho^3\right)(P - R) \cdot T + 1/\rho^2. \quad (14)$$

At s_0 we have

$$P - R = 0;$$

thus, (14) becomes $1/\rho^2$ and

$$u(P) \sim A(s_0)(\exp iks_0) \cdot \int_{-\infty}^{\infty} \exp\{ik(s - s_0)^3/6\rho^2(s_0)\} ds/\rho(s). \quad (15)$$

Proceeding as before, we obtain from (15)

$$u(P) \sim A(s_0)(\exp iks_0) \int_{-\infty}^{\infty} \exp\{ik\rho(s_0)\theta^3/6\} d\theta \quad (16)$$

or

$$u(P) \sim \frac{2\Gamma(\frac{1}{3}) \cos \pi/6}{3\{k\rho(s_0)/6\}^{1/3}} A(s_0) \exp iks_0.$$

(See Appendix I.⁵)

If the caustic C is "flat" (*i.e.*, $\rho = \infty$) at the point s_0 , so that the third derivative (14) vanishes, we must use the next order term in the power series expansion of the phase in (3). For the general case in which the first n terms of the power series vanish, we use the $(n+1)$ st term, and the procedure is the same as that for the more regular cases just described.

Where the caustic C has a cusp, the radius vector $R(s)$ has a singularity. It is convenient to assume that the origin of our coordinate system is at the cusp which we can take to be the point $s=0$ on C . Then the cusp is at $R(0)=0$. We can also assume that the coordinate system is so oriented that the tangent vector $T(s)$ has a horizontal limiting position at $s=0$; *i.e.*, at $s=0$

$$T(0) = (1, 0).$$

Geometrically, a cusp is a point at which two branches of the curve C meet. The field should be computed separately for each branch and the total field obtained by superposing the two contributions. Thus, we now consider the contribution of a single branch for which s ranges from zero through all positive values.

It is appropriate to assume that there is a singularity in the curvature $1/\rho(s)$ at the cusp and we shall assume that the singularity is algebraic:

$$1/\rho(s) = \lambda s^\alpha + \mu s^\beta + \text{a function of higher order in } s. \quad (17)$$

Here $\beta > \alpha$ and we also assume that $\alpha > -1$, which is necessary in order that $T(s)$ exist at $s=0$.

Now we can write the following for the components of $R(s)$ in general:

$$R(s) = \left\{ \int_0^s \left[\cos \int_0^\sigma dt/\rho(t) \right] d\sigma, \int_0^s \left[\sin \int_0^\sigma dt/\rho(t) \right] d\sigma \right\}, \quad (18)$$

⁵ This result agrees with that of Landau and Lifshitz, *op. cit.*, p. 160.

from which

$$T(s) = \left\{ \cos \int_0^s dt/\rho(t), \sin \int_0^s dt/\rho(t) \right\}, \quad (19)$$

and

$$\begin{aligned} \frac{dT}{ds} &= N(s)/\rho(s) \\ &= \left\{ -\sin \int_0^s dt/\rho(t), \cos \int_0^s dt/\rho(t) \right\} / \rho(s). \end{aligned} \quad (20)$$

From (18) and (19) using the power series expansions for the sin and cos functions about zero, we obtain

$$\begin{aligned} R &= \left\{ s - [\lambda^2/2(\alpha+1)^2(2\alpha+3)]s^{2\alpha+3} - \dots, \right. \\ &\quad \left. [\lambda/(\alpha+1)(\alpha+2)]s^{\alpha+2} \right. \\ &\quad \left. + [\mu/(\beta+1)(\beta+2)]s^{\beta+2} + \dots \right\} \end{aligned} \quad (21)$$

$$\begin{aligned} T &= \left\{ 1 - \frac{1}{2}[\lambda/(\alpha+1)]s^{\alpha+1} + [\mu/(\beta+1)]s^{\beta+1} \right\}^2 + \dots, \\ &\quad \left\{ [\lambda/(\alpha+1)]s^{\alpha+1} + [\mu/(\beta+1)]s^{\beta+1} + \dots \right\}. \end{aligned} \quad (22)$$

Writing $P = (x, y)$ we then have for the phase,

$$\begin{aligned} P - R \cdot T + s &= x + [y\lambda/(\alpha+1)]s^{\alpha+1} + [y\mu/(\beta+1)]s^{\beta+1} \\ &\quad - [x\lambda^2/2(\alpha+1)^2]s^{2\alpha+2} - [x\mu^2/2(\beta+1)^2]s^{2\beta+2} \\ &\quad - [x\lambda\mu/(\alpha+1)(\beta+1)]s^{\alpha+\beta+2} \\ &\quad + [\lambda^2/(\alpha+2)(2\alpha+3)]s^{2\alpha+3} + \dots \end{aligned} \quad (23)$$

When P is at the cusp, $x=y=0$, and the phase becomes

$$[\lambda^2/(\alpha+2)(\alpha+3)]s^{2\alpha+3} + \dots \quad (24)$$

Along a ray, $y=0$, and the phase becomes near the cusp

$$x - [x\lambda^2/2(\alpha+1)^2]s^{2\alpha+2} + \dots \quad (25)$$

We can also consider various other cases in which we approach the cusp along different paths, *e.g.*, along $x=0$, and in each case we take from (23) only the lowest order term in s . For these other approaches a more precise relationship between α and β must be given to decide which term is actually of lowest order.

In (3) we can replace $A(s)$ by $A(0)$ and the phase by (24), (25) or by a special case of (23), depending on the particular limiting case that we wish to compute. Consider, for example, the case of $x=y=0$. From (3), (24) and the Appendix we have

$$\begin{aligned} u(P) &\sim A(0)(\operatorname{sgn} \lambda) \left| \lambda \right|^{1/(2\alpha+3)} (2\alpha+3)^{-(\alpha+2)/(2\alpha+3)} \\ &\quad \cdot k^{-(\alpha+1)/(2\alpha+3)} (\alpha+2)^{(\alpha+1)/(2\alpha+3)} \Gamma[(\alpha+1)/(2\alpha+3)] \\ &\quad \cdot \exp \{ i\pi(\alpha+1)/2(2\alpha+3) \}. \end{aligned} \quad (26)$$

The contribution of the other branch of the cusp in the case of complete symmetry about the horizontal is calculated in the same way. However, we must be careful that the integration is in such a direction that $T(s)$ turns so that its angle with the horizontal

$$\theta = \int_0^s ds/\rho(s)$$

increases. On the other hand, if we assume that the second branch is also given for s , going from zero through positive values, it is clear geometrically that $\rho(s)$ must have a sign opposite that of $\rho(s)$ on the first branch. The integration over C for the second branch must then be taken in the direction, $-ds$. Now to first order in s we have from (17),

$$\operatorname{sgn} \lambda = \operatorname{sgn} \rho(s).$$

Thus, in the case $x=y=0$, an examination of (26) indicates that a sign change occurs for integration in the direction ds , and, therefore, integration in the direction $-ds$ exactly cancels the effect of the sign change. The contribution of the second branch will then be exactly the same as that of the first branch, and the total field at the cusp will be exactly twice the value given by (26) with $\operatorname{sgn} \lambda$ equal to $+$.

It is interesting to consider also the case where P is on the ray $y=0$ through the cusp. From (3), (25) and the Appendix,

$$\begin{aligned} u(P) &\sim (\operatorname{sgn} \lambda) A(0) (\pi/2k |x|)^{1/2} \\ &\quad \cdot \exp \{ ikx - (i\pi/4) \operatorname{sgn} x \}. \end{aligned} \quad (27)$$

The value of $u(P)$ given by (27) when $\operatorname{sgn} \lambda$ is $+$ is exactly one-half that given by (8) for a point P on an ordinary ray in the lit region. This factor of one-half occurs because we are considering just the effect of one branch of the cusp. The other branch of the cusp produces a contribution to $u(P)$ identical with that given by (27). The sum of the two contributions will then result in a field value for $u(P)$ which is the same as that occurring for an ordinary position of P in a lit region.

For the case of P on the line, $x=0$ orthogonal to the ray $y=0$, we choose the second and third terms on the right side of (23) for the phase:

$$[y\lambda/(\alpha+1)]s^{\alpha+1} + [y\mu/(\beta+1)]s^{\beta+1}. \quad (28)$$

From (3) and (28) we have

$$\begin{aligned} u(P) &\sim \lambda A(0) \int_0^\infty s^\alpha \exp ik \\ &\quad \cdot \{ [y\lambda/(\alpha+1)]s^{\alpha+1} + [y\mu/(\beta+1)]s^{\beta+1} \} ds \\ &= [\lambda A(0)/(\alpha+1)] \int_0^\infty \exp ik \\ &\quad \cdot \{ [y\lambda/(\alpha+1)]\tau + [y\mu/(\beta+1)]\tau^{(\beta+1)(\alpha+1)} \} d\tau. \end{aligned} \quad (29)$$

For certain relations between α and β , *e.g.*, for $(\beta+1)/(\alpha+1)$ an integer, (29) can be evaluated easily by stationary phase. For the contribution of the other branch of the cusp, we multiply the result (29) by -1 and replace λ and μ by $-\lambda$ and $-\mu$.

IV. AN EXAMPLE: A CIRCULAR CAUSTIC AND A POINT FOCUS

Let us consider a simple example of a caustic with no cusp. Let C be a circle of radius ρ with its center at the

origin. Then C is given by

$$\mathbf{R}(s) = (\rho \cos(s/\rho), \rho \sin(s/\rho)). \quad (30)$$

We also have

$$\begin{aligned} \mathbf{T} &= (-\sin(s/\rho), \cos(s/\rho)) \\ \mathbf{N} &= -(\cos(s/\rho), \sin(s/\rho)), \end{aligned}$$

and ρ is the constant radius of curvature of C . The relation (3) becomes

$$\begin{aligned} u(\mathbf{P}) &= \int_C A(s) \exp \{ ik[-x \sin(s/\rho) \\ &\quad + y \cos(s/\rho) + s] \} ds / \rho \\ &= \int_C B(\theta) \exp \{ ik[r \sin(\phi - \theta) + \rho\theta] \} d\theta, \end{aligned} \quad (31)$$

where $A[s(\theta)] = B(\theta)$, $r = |\mathbf{P}|$ and ϕ is the direction angle of \mathbf{P} , and $s = \rho\theta$. The angle θ_0 for which $\mathbf{P} - \mathbf{R}$ is normal to \mathbf{N} can be computed quite easily or seen at once geometrically

$$\theta_0 = \phi - \cos^{-1}(\rho/r). \quad (32)$$

The distance D_0 is easily seen to be

$$D_0 = (r^2 - \rho^2)^{1/2}. \quad (33)$$

From (8) for a point \mathbf{P} in the lit region $r > \rho$ we have

$$\begin{aligned} u(\mathbf{P}) &\sim \{ 2\pi/k(r^2 - \rho^2)^{1/2} \}^{1/2} B(\theta_0) \\ &\cdot \exp \{ ik[(r^2 - \rho^2)^{1/2} + \rho(\phi - \cos^{-1}(\rho/r)) - i\pi/4] \} \end{aligned} \quad (34)$$

where we have used $\rho\theta_0 = s_0$ and (32).

The result (34) holds only in the case when a single portion of the wavefront corresponding to C passes through \mathbf{P} . In general, a spiral-shaped wavefront winds around C and, as it propagates, one turn of the spiral after another will pass through \mathbf{P} . At each passage the stationary point is such that the corresponding angle θ_0 is increased by 2π . This fact is predictable by the observation that the term $\cos^{-1}(\rho/r)$ of (32) is ambiguous. Thus, in general, the field at \mathbf{P} is

$$\begin{aligned} u(\mathbf{P}) &\sim \sum_{n=-\infty}^{\infty} B(\theta_0 + 2\pi n) \{ 2\pi/k(r^2 - \rho^2)^{1/2} \}^{1/2} \\ &\cdot \exp \{ ik[(r^2 - \rho^2)^{1/2} + \rho(\phi - \cos^{-1}(\rho/r) - 2\pi n)] \}. \end{aligned} \quad (35)$$

Of course, $B(\theta + 2\pi n)$, which is given by the source, may vanish for an argument sufficiently large in magnitude and reduce (35) to a finite sum.

For \mathbf{P} on C where $r = \rho$ we have from (16)

$$\begin{aligned} u(\mathbf{P}) &\sim \{ 2\Gamma(1/3)(\cos \pi/6)/3[k\rho/6]^{1/3} \} A(\rho\phi) \\ &\cdot \exp(ik\rho\phi) \end{aligned} \quad (36)$$

for a single turn of the spiral wavefront. In general,

$$\begin{aligned} u(\mathbf{P}) &\sim \{ 2\Gamma(1/3)(\cos \pi/6)/3[k\rho/6]^{1/3} \} \\ &\cdot \sum_{n=-\infty}^{\infty} A(\rho\phi + 2n\pi\rho) \exp[ik(\rho\phi + 2\pi n)]. \end{aligned} \quad (37)$$

In the case of a point focus, the radius ρ of C becomes zero, and we might consider this as a limiting case of a circular caustic. Setting $\rho = 0$ in (31) we obtain

$$u(\mathbf{P}) \sim \int_C B(\theta) \exp \{ ik[r \sin(\phi - \theta)] \} d\theta. \quad (38)$$

Now in the limit $r \rightarrow 0$, a stationary point θ_0 will lag behind ϕ by $\pi/2$. It is convenient to replace the function $B(\theta)$ in (38) by an angular source distribution function which determines the proportion of energy assigned to each ray. Since the direction of a ray through \mathbf{P} is ϕ , this can be accomplished by a change of variable,

$$\theta + \pi/2 = \xi. \quad (39)$$

Correspondingly we set

$$F(\xi) = B(\xi - \pi/2). \quad (40)$$

Then (38) becomes

$$u(\mathbf{P}) = \int_{-\pi}^{\pi} F(\theta) \exp \{ ikr \cos(\phi - \theta) \} d\theta, \quad (41)$$

where we have assumed only a single integration about C which implies that $F(\theta)$ is periodic of period 2π . If this were not true, we should be forced to extend the integral (41) over a larger interval. We can check on the meaning of $F(\theta)$ with respect to our source by evaluating (41) directly by stationary phase:

$$u(\mathbf{P}) \sim (2\pi/kr)^{1/2} F(\phi) \exp \{ ikr + i\pi/4 \}, \quad (42)$$

which holds when \mathbf{P} is not near the focus.

The result (41) agrees with the Debye solution of the perfect focus problem.³

In case ρ is not zero and the point \mathbf{P} is inside the circular caustic, we have

$$r < \rho, \quad (43)$$

and the stationary point given by (32) is complex. Since

$$\cos(\phi - \theta_0) = \frac{\rho}{r},$$

it is easily found in the complex case that

$$\theta_0 = \phi + i \log \{ (\rho - \sqrt{\rho^2 - r^2})/r \}. \quad (44)$$

Then (34) becomes in this case

$$\begin{aligned} u(\mathbf{P}) &\sim [2\pi/k(\rho^2 - r^2)^{1/2}]^{1/2} B(\theta_0) \\ &\cdot \exp \{ -k[(\rho^2 - r^2)^{1/2} + \rho \log(\rho/r - \sqrt{\rho^2 - r^2}/r)] \\ &\quad + ik\phi \}, \end{aligned} \quad (45)$$

where $B(\theta_0)$ is the analytic continuation of the source distribution to the complex angle θ_0 given by (44).

V. CONCLUSION

It can be seen from an examination of the various formulas for the field amplitudes given in Sections III and IV that the field is completely specified in the small

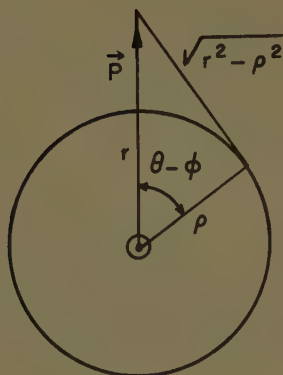


Fig. 2.

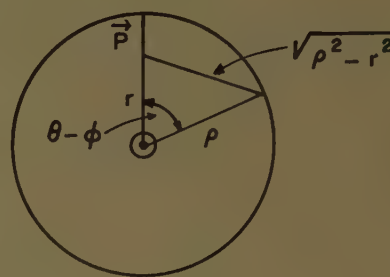


Fig. 3.

wavelength limit by the local geometry determined by the caustics of the ray system. For points in the lit region or on a smooth part of a caustic, it is necessary to know the local geometry only to the order of the radius of curvature of the caustic. In the neighborhood of a cusp of a caustic, the necessary information must include the order of contact of the two caustic branches. In the shadow region an approximate value for the field amplitude near the caustic is again given in terms of the curvature of the caustic. For points further away in the shadow region, the small wavelength limit of the field amplitude is exponentially small, and it can be argued that nonzero values for this case occur only when the wavelength is sufficiently long to introduce nongeometrical considerations into the problem. In case a point in the shadow region is near enough to several points of closest approach on a caustic, the field should consist of a superposition of fields determined by (45) for each such point of closest approach. (See Figs. 2 and 3.)

APPENDIX

For the sake of completeness we consider the evaluation of an integral of the form

$$I = \int_0^{\infty} \exp \{ i k a s^{\lambda} \} ds, \quad (46)$$

where a is a real constant and λ is a real positive constant.

After a change of variable

$$s = (|a|k)^{-1/\lambda} \tau^{1/\lambda} \exp \{ (\pi/2\lambda) \operatorname{sgn} a \} \quad (47)$$

we obtain, in (46),

$$\begin{aligned} I &= \{ (|a|k)^{-1/\lambda} / \lambda \} \{ \exp [(\pi/2\lambda) \operatorname{sgn} a] \} \\ &\quad \cdot \int_0^{-i\infty} \tau^{1/\lambda-1} \exp(-\tau) d\tau \\ &= \{ (|a|k)^{-1/\lambda} / \lambda \} \{ \exp [(\pi/2\lambda) \operatorname{sgn} a] \} \Gamma\left(\frac{1}{\lambda}\right). \end{aligned} \quad (48)$$

On the Discontinuity Problem at the Input to an Anisotropic Waveguide*

A. D. BRESLER†

Summary—The problem considered is that of the discontinuity at the transverse plane separating two regions of a dissipationless uniform waveguide, one isotropic, the other anisotropic. The integral equation for this discontinuity is formulated in terms of the "four-vector" guides modes which propagate in the two regions. Variational expressions based on this integral equation are then given for the scattering coefficients which describe the effect of the junction on the propagating modes of the two regions. These variational expressions are then employed to obtain numerical results for the scattering coefficients at the transverse plane separating two regions of a rectangular waveguide, one empty, the other filled with a transversely magnetized dissipationless ferrite. These numerical results are found to agree reasonably well with those obtained from an alternative solution to this problem given by Sharpe and Heim.

INTRODUCTION

MOST of the published work on anisotropic discontinuities has been concerned with scattering by small anisotropic discontinuities introduced into isotropic waveguides. The following constitutes a representative list of the published work in this area: Berk and Lengyel¹ and Berk and Strumwasser² obtain the quasi-static solution for scattering by small ferrite spheres and thin ferrite rods. Berk and Strumwasser,² Damon,³ and Stinson⁴ study the coupling of isotropic waveguides by ferrite loaded apertures. Hurd^{5,6} expresses the solution for the scattering of a plane wave by a small anisotropic ellipsoid as the quasi-static solution plus higher-order corrections and develops expressions for the first few corrections.

Very little attention has been paid to problems in which the anisotropic discontinuity is not small. Tyras and Held⁷ have considered the radiation from a ferrite-filled aperture. However, in formulating their solution they ignored the discontinuity at the junction of the empty and ferrite-filled waveguides and they obtained the radiation pattern by assuming the aperture illumi-

nation to be identical with the incident waveguide fields. Van Trier⁸ and Gintsburg⁹ give brief discussions of the discontinuity at the interface separating two regions of a rectangular waveguide, one empty and the other completely filled with a transversely magnetized ferrite. These discussions do not go very much beyond pointing out that a single mode incident on the junction excites an infinity of modes in both the empty and ferrite loaded regions.¹⁰ Epstein¹¹ formulates the solution to this problem in terms of an infinite system of simultaneous linear equations and employs a method of successive approximations (the convergence of which is not assured) to obtain a power-series solution for the input reflection coefficient on the empty waveguide side. Sharpe and Heim¹² formulate the solution to this same problem in terms of an integral equation for the electric field in the junction plane. A convergent iteration procedure is then employed to obtain the solution to this integral equation for the special case where both the empty and ferrite-filled rectangular waveguides support only a single pair of propagating modes.

In this paper, we will consider the discontinuity problem at the transverse plane (junction plane) which divides a passive dissipationless infinite uniform waveguide into two regions, one isotropic, the other anisotropic. (See Fig. 1.) The discussion will proceed on the

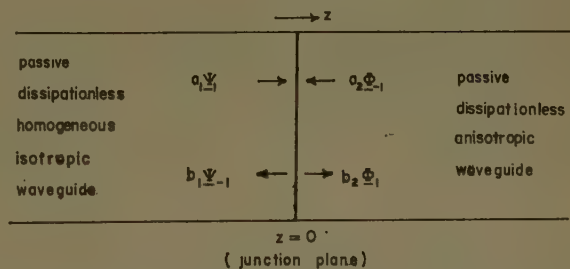


Fig. 1—Illustrating the discontinuity problem under consideration.

assumption that each of the two waveguides supports only a single pair of propagating modes. The modifications required to take into account additional propagat-

* The investigation reported herein has been sponsored by the AF Cambridge Research Center, under Contract No. AF-19(604)-2031.

† Jasik Laboratories, Westbury, N. Y. Formerly with the Microwave Res. Inst., Polytechnic Institute of Brooklyn, N. Y.

¹ A. D. Berk and B. A. Lengyel, "Magnetic fields in small ferrite bodies with applications to microwave capacities containing such bodies," *Proc. IRE*, vol. 43, pp. 1587-1591; November, 1955.

² A. D. Berk and E. Strumwasser, "Ferrite directional couplers," *Proc. IRE*, vol. 44, pp. 1439-1445; October, 1956.

³ R. W. Damon, "Magnetically controlled microwave direction coupler," *J. Appl. Phys.*, vol. 26, pp. 1281-1283; October, 1955.

⁴ D. C. Stinson, "Coupling through an aperture containing an anisotropic ferrite," *IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES*, vol. MTT-5, pp. 184-191; July, 1957.

⁵ R. A. Hurd, "Scattering from a small anisotropic ellipsoid," *Can. J. Phys.*, vol. 36, pp. 1058-1071; 1958.

⁶ —, "The magnetic fields of a ferrite ellipsoid," *Can. J. Phys.*, vol. 36, pp. 1072-1083; 1958.

⁷ G. Tyras and G. Held, "Radiation from a rectangular waveguide filled with ferrite," *IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES*, vol. MTT-6, pp. 268-277; July, 1958.

⁸ A. A. van Trier, "Guided electromagnetic waves in anisotropic media," *Appl. Sci. Res., B*, vol. 3, pp. 305-371; 1953.

⁹ M. A. Gintsburg, "On waves in a gyrotropic medium," *Bull. Acad. Sci., USSR, Phys. Ser.*, vol. 18, no. 4, pp. 444-456; 1954.

¹⁰ This is in marked contrast with the analogous problem for a dielectric-filled waveguide where the solution to the boundary value problem at the interface involves only those modes for which the dependence of the transverse fields on the transverse coordinates is identical to that of the incident mode.

¹¹ P. S. Epstein, "Theory of wave propagation in a gyromagnetic medium," *Rev. Mod. Phys.*, vol. 28, pp. 3-17; January, 1956.

¹² C. B. Sharpe and D. S. Heim, "A ferrite boundary value problem in rectangular waveguide," *IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES*, vol. MTT-6, pp. 42-46; January, 1958.

ing modes are considered elsewhere.¹³ The solution to the discontinuity problem will be obtained by formulating an integral equation which expresses the continuity of the transverse fields through the junction plane. To formulate this integral equation it will be necessary to express the fields in the two waveguides in terms of the "four-vector" guided modes of these waveguides.¹⁴ This requirement follows from the fact that, in general, the modes of an anisotropic waveguide cannot be described by a statement of only either the transverse electric field or the transverse magnetic field; to specify one of these modes requires a statement of both these fields where, in general, these are not simply related.

The identification of the four-vector guided modes as the proper modes (eigenfunctions) for an arbitrary uniform waveguide has already been made.¹⁴ This identification is based on an operator (matrix) formulation for the Maxwell equations which was employed by Bresler *et al.*,¹⁴ to establish the orthogonality properties of these modes, and again by the author¹⁵ to obtain information on the manner in which the modes contribute to the fields excited by a point source (*i.e.*, to the Green's function). More extensive discussions of the operator formulation for the Maxwell equations are also given in the literature.^{13,16,17} Vector formulations for the Maxwell equations based on (and equivalent to representations of) the operator formulation are given by the author,^{18,19} where the later work gives a concise summary of the contents of the earlier. Since many of the references cited in this paragraph are not readily available, we will attempt to render the discussion in this paper as self-contained as possible, consistent with the requirement of reasonable length. With this end in view, the discussion will begin with a consideration of the four-vector modes and of their orthogonality properties. These will then be employed to exhibit the completeness relations (*idem* operator representations) which reflect the assumed completeness of these mode sets. These completeness relations, together with results previously

established,¹⁵ will then enable us to obtain expressions for the Green's functions appropriate to the two waveguides.

Once the Green's functions for the waveguides are available, it is a simple matter to obtain the integral equation which expresses the continuity of the transverse fields through the junction plane. Variational expressions based on this integral equation will then be given for the scattering coefficients which describe the effect of the junction on the propagating modes of the two regions. These variational expressions will then be employed to obtain numerical results for the scattering coefficients at the junction of two rectangular waveguides, one empty, the other filled with a transversely magnetized ferrite. These numerical results will then be compared with those obtained from the alternative solution to this problem given by Sharpe and Heim.¹²

THE MODES AND THEIR ORTHOGONALITY PROPERTIES

The modes of the isotropic waveguide will be designated by Ψ_n , while those of the anisotropic waveguide will be designated by Φ_m . These modes are identified with the indicated column matrices²⁰

$$\Phi_n \rightarrow \begin{pmatrix} E_{tn} \\ iH_{tn} \end{pmatrix} \quad \Phi_m \rightarrow \begin{pmatrix} E_{tm} \\ iH_{tm} \end{pmatrix}, \quad (1)$$

where E_t and H_t are, respectively, the transverse electric and magnetic fields of the mode.¹⁴ It is assumed that the dependence on time, t , and on the axial coordinate of the waveguide, z , for all modes is given by $\exp i(\kappa z - \omega t)$ where ω is the radian frequency and κ the propagation constant along z . The indicated dependence on t and z is suppressed so that Ψ_n and Φ_m are functions of the transverse coordinates only. These mode functions are obtained as solutions to the eigenvalue problems:^{13,14}

$$(L_A - \kappa_{A,n} \Gamma_s) \Psi_n = 0$$

$$(L_F - \kappa_{F,m} \Gamma_s) \Phi_m = 0, \quad (2)$$

where A and F distinguish quantities associated with the isotropic and anisotropic waveguides, respectively. The "transverse Maxwell operators" L_A and L_F contain within them all the information concerning the media filling the two waveguides and also all the differentiation operations appropriate to the transverse field equations. For our purposes it is not necessary to exhibit representations for these operators. The interested reader will find these in the literature.^{13,14} We remark that, as in the case of the representation for Γ_s exhibited below, these representations take the form of 2-by-2 matrices with dyadic elements. The rules for the operation of such

¹³ A. D. Bresler, "On the Discontinuity Problem at the Input to an Anisotropic Waveguide," D.E.E. dissertation, Polytechnic Inst. of Brooklyn, N. Y., June, 1959; to be issued by their Microwave Res. Inst. as Res. Rept. No. R-716-59.

¹⁴ A. D. Bresler, G. H. Joshi, and N. Marcuvitz, "Orthogonality properties for modes in passive and active uniform waveguides," *J. Appl. Phys.*, vol. 29, pp. 794-799; May, 1958.

¹⁵ A. D. Bresler, "The far fields excited by a point source in a passive dissipationless anisotropic uniform waveguide," *IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES*, vol. MTT-7, pp. 282-287; April, 1959.

¹⁶ A. D. Bresler and N. Marcuvitz, "Operator Methods in Electromagnetic Field Theory; Part I—Abstract Operator Formulations for the Maxwell Equations," *Microwave Res. Inst., Polytechnic Inst. of Brooklyn, N. Y., Res. Rept. No. R-495-56*; May, 1956.

¹⁷ ———, "Operator Methods in Electromagnetic Field Theory; Part II—Guided Modes in Uniform Cylindrical Waveguide Regions," *Microwave Res. Inst., Polytechnic Inst. of Brooklyn, N. Y., Res. Rept. No. R-565-57*; March, 1957.

¹⁸ A. D. Bresler, "Vector Formulations for the Electromagnetic Field Equations in Uniform Waveguides Containing Anisotropic Media," *Microwave Res. Inst., Polytechnic Inst. of Brooklyn, N. Y., Res. Rept. No. R-676-58*; September, 1958.

¹⁹ A. D. Bresler, "Vector Formulations for the field equations in anisotropic waveguides," *IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES*, vol. MTT-7, p. 298; April, 1959.

²⁰ The arrows used in, *e.g.*, (1) and (3) indicate that the matrices constitute representations of abstract elements and operators in a properly defined space.

operators on elements (column matrices), or for the sequential operation of two operators, include the normal rules of matrix algebra with the understanding that the dot product is implied for products of dyadics and vectors or of two dyadics. The operator Γ_z is defined by²⁰

$$\Gamma_z \rightarrow \begin{pmatrix} 0 & iz_0 \times 1_t \\ iz_0 \times 1_t & 0 \end{pmatrix}, \quad (3)$$

where z_0 is the unit vector in the z direction and 1_t is the unit dyadic in the plane transverse to z . It is of interest to note that

$$\Gamma_z^2 = I_t \rightarrow \begin{pmatrix} 1_t & 0 \\ 0 & 1_t \end{pmatrix}, \quad (4)$$

where I_t plays the role of the unit operator in the space of all elements of the type indicated in (1).

In the discussion which follows, we shall make extensive use of the hermitian inner product of two elements defined thus:

$$(\Psi_\beta, \Psi_\alpha) = \int_S \int [E_{i\beta}^* \cdot E_{t\alpha} + (iH_{i\beta})^* \cdot (iH_{t\alpha})] dS, \quad (5)$$

where the asterisk indicates the complex conjugate. This inner-product definition requires an integration over the cross section, S , of the waveguide in addition to the operations usually required to obtain the hermitian inner product of column matrices. Based on this inner product there is associated with a given operator, e.g., L_A , an adjoint operator L_A^+ via the adjointness relation

$$(\Psi^+, L_A \Psi) = (L_A^+ \Psi^+, \Psi) = 0. \quad (6)$$

When there are differentiation operations implicit in an operator (as is the case with L_A), this provides only a formal definition for the adjoint. The complete definition requires a statement of the boundary conditions to be satisfied by the elements on which the operator and its adjoint are to operate.¹³⁻¹⁵ This additional requirement will not concern us here.

It is readily verified that the operator Γ_z is hermitian, i.e., that $\Gamma_z = \Gamma_z^+$. Since both the isotropic and anisotropic waveguides are dissipationless it follows that $L_A = L_A^+$ and $L_F = L_F^+$.^{14,15} Also, the modal propagation constants of both waveguides will occur in complex conjugate pairs.¹² Since the isotropic waveguide is reflection symmetric, its modal propagation constants must also occur in positive and negative pairs.¹⁷ Further, since the waveguide to the left of $z=0$ in Fig. 1 has been assumed to be a homogeneous isotropic waveguide, we are assured that all the propagation constants of this waveguide are either real (for the propagating modes) or imaginary (for the cutoff modes).¹⁷ We now associate the modes $\Psi_{\pm n}$ with the modal propagation constants $\kappa_{A,\pm n} = \pm \kappa_{A,n}$. Recalling that we have assumed that the isotropic waveguide supports only a single pair of propagating modes, we recognize that $\pm \kappa_{A,1}$ are real, while $\pm \kappa_{A,n}$, $n > 1$, are imaginary and adopt the convention that $\kappa_{A,1} > 0$, $\text{Im } \kappa_{A,n} > 0$ for $n > 1$. Finally, we

assume that, as is always possible,¹⁷ the modes of the homogeneous isotropic waveguide have been chosen so that

$$\begin{aligned} \Psi_{-n} &= \Gamma_z \Psi_n & (\text{for all } n) \\ \Psi_{\pm 1}^+ &= \Psi_{\pm 1}^* & \Psi_{\pm 1}^* = \Psi_{\mp 1} \\ \Psi_{\pm n}^+ &= \Psi_{\mp n}^* & \Psi_{\pm n}^* = \Psi_{\mp n} \quad (n > 1), \end{aligned} \quad (7)$$

where

$$\Gamma_z \rightarrow \begin{pmatrix} 1_t & 0 \\ 0 & -1_t \end{pmatrix} \quad (8)$$

and Ψ^+ is the adjoint eigenfunction (mode) whose identification is required in the statement of the orthogonality properties of the modes:¹⁴

$$(\Psi_\beta^+, \Gamma_z \Psi_\alpha) = N_\alpha \delta_{\alpha\beta} \quad (9)$$

with $\delta_{\alpha\beta} = 1$ for $\kappa_\alpha = \kappa_\beta$ and $\delta_{\alpha\beta} = 0$ for $\kappa_\alpha \neq \kappa_\beta$. In (9), N_α is an arbitrary normalization constant. It then follows from (7) and (9) that

$$\begin{aligned} (\Psi_{\pm 1}, \Gamma_z \Psi_{\pm 1}) &= \pm N_1 & N_1^* &= N_1 > 0 \\ (\Psi_{\mp n}, \Gamma_z \Psi_{\pm n}) &= \pm N_n & N_n^* &= -N_n = N_{-n} \quad (n > 1) \end{aligned} \quad (10)$$

with all other inner products involving these modes equal to zero.

The modal propagation constants of the dissipationless anisotropic waveguide are, in general, restricted only by the requirement that they occur in complex conjugate pairs. Thus, there is no implication that the propagation constants for the two propagating modes of this waveguide, $\kappa_{F,1}$ and $\kappa_{F,-1}$ are in any way related. Further, the mode functions Φ_1 and Φ_{-1} are not simply related. Similarly, in general, there is no requirement for a simple relationship between the mode functions Φ_m and Φ_m^* associated with the (in general, complex) cutoff mode ($m > 1$) propagation constants $\kappa_{F,m}$ ($\text{Im } \kappa_{F,m} > 0$) and $\kappa_{F,m}^* = \kappa_{F,m}^*$, respectively. Therefore, in general, we cannot make statements for the modes of the anisotropic waveguide similar to the reflection symmetry and conjugation properties in (7). The identifications of the adjoint eigenfunctions are

$$\Phi_{\pm 1}^+ = \Phi_{\pm 1}; \quad \Phi_m^+ = \Phi_m^* \quad m \neq 1. \quad (11)$$

It then follows from (9) that

$$\begin{aligned} (\Phi_{\pm 1}, \Gamma_z \Phi_{\pm 1}) &= M_{\pm 1} \begin{cases} M_1^* = M_1 > 0 \\ M_{-1}^* = M_{-1} < 0 \end{cases} \\ (\Phi_m^*, \Gamma_z \Phi_m) &= M_m \quad M_m^* = M_m^*, \end{aligned} \quad (12)$$

with all other inner products involving these modes equal to zero.

At this point it is pertinent to digress briefly and remark on the significance of the restrictions $M_1 > 0$, $M_{-1} < 0$. First, we note that, in a dissipationless waveguide, the normalization constant associated with a propagating mode equals twice the net real power flow (along $+z$) associated with that mode.¹⁵ Thus we see that we have assumed here that one of the two propa-

gating modes of the anisotropic waveguide carries power along $+z$ while the other propagating mode carries power along $-z$. In an anisotropic waveguide there need be no correlation between the direction of power flow associated with a propagating mode and its direction of (phase) propagation. Thus, the arrows shown in Fig. 1 indicate the directions of power flow associated with the corresponding modes and not necessarily the directions of (phase) propagation. In an earlier paper,¹⁵ the author has shown that a propagating mode of any passive dissipationless uniform waveguide contributes to the field excited by a source introduced into this waveguide only in that direction in which it transports energy away from the source. Therefore, since $N_1 > 0$ and $M_{-1} < 0$, the modes Ψ_1 and Φ_{-1} can be excited by sources at $z \rightarrow -\infty$ and $z \rightarrow +\infty$, respectively. Further, since $N_{-1} < 0$ and $M_1 > 0$, the modes Ψ_{-1} and Φ_1 may be excited by the equivalent induced sources in the junction plane.

THE COMPLETENESS RELATIONS

We assume that each of the two sets of modes, Ψ_n and Φ_m , are complete in the space of all transverse elements of the type in (1) which are defined in the cross section, S , of the waveguide. This assumption implies the existence of appropriate completeness relations (representations of the idem operator I_t) in terms of each of these two sets. Because Γ_z plays the role of a weight function (operator) in the eigenvalue problems (2) it is more convenient first to exhibit representations for this operator. In the discussion which follows, we will state all results in terms of the set Φ_m recognizing that, in each instance, an analogous statement exists in terms of the set Ψ_n . The representation given below for Γ_z exploits the fact that the modes Φ_m are eigenfunctions (with respect to the weight operator Γ_z) of a hermitian operator so that the adjointness identification in (11) obtains.

It is shown¹⁷ that the following constitute alternative representations for Γ_z in terms of the set Φ_m :

$$\Gamma_z = \sum_{\alpha} \frac{1}{M_{\alpha}} \Phi_{\alpha} \Phi_{\alpha}^* = \sum_{\alpha} \frac{1}{M_{\alpha}^*} \Phi_{\alpha}^* \Phi_{\alpha} \quad (13)$$

where the sum over α implies the sum of contributions from all the eigenvalues $\kappa_{F,\alpha}$ of L_F with the number of contributions for each eigenvalue equal to the multiplicity (degeneracy) of that eigenvalue. Each term of either sum, e.g.,

$$\frac{1}{M_{\alpha}} \Phi_{\alpha} \Phi_{\alpha}^* \rightarrow \frac{1}{M_{\alpha}} \begin{pmatrix} E_{t\alpha} E_{t\alpha}^* & -i E_{t\alpha} H_{t\alpha}^* \\ i H_{t\alpha} E_{t\alpha}^* & H_{t\alpha} H_{t\alpha}^* \end{pmatrix}, \quad (14)$$

is an operator dyad with the rule for its operation given by

$$\left(\xi, \left[\frac{1}{M_{\alpha}} \Phi_{\alpha} \Phi_{\alpha}^* \right] \cdot \mathbf{n} \right) = \frac{1}{M_{\alpha}} (\xi, \Phi_{\alpha}) (\Phi_{\alpha}^*, \mathbf{n}), \quad (15)$$

It follows from this definition that

$$\left[\frac{1}{M_{\alpha}} \Phi_{\alpha} \Phi_{\alpha}^* \right]^+ = \frac{1}{M_{\alpha}^*} \Phi_{\alpha}^* \Phi_{\alpha}, \quad (16)$$

so that (13) expresses the fact that $\Gamma_z^+ = \Gamma_z$.

Since $\Gamma_z^2 = I_t$, the representations (13) are equivalent to a variety of representations for I_t . Before exhibiting these, we point out that to be consistent with the definition in (15) it is necessary to define operation from the right on an element by an operator as equivalent to operation from the left on the same element by the adjoint of the operator. Thus, if we operate on (13) with Γ_z acting first from the left and then from the right, we obtain the following alternative representations for I_t :

$$\begin{aligned} I_t &= \sum_{\alpha} \frac{1}{M_{\alpha}} \Gamma_z \Phi_{\alpha} \Phi_{\alpha}^* = \sum_{\alpha} \frac{1}{M_{\alpha}} \Phi_{\alpha} \Gamma_z \Phi_{\alpha}^* \\ &= \sum_{\alpha} \frac{1}{M_{\alpha}^*} \Gamma_z \Phi_{\alpha}^* \Phi_{\alpha} = \sum_{\alpha} \frac{1}{M_{\alpha}^*} \Phi_{\alpha}^* \Gamma_z \Phi_{\alpha}. \end{aligned} \quad (17)$$

To avoid misinterpretation, we specify that when Γ_z appears between the two elements of a dyad it operates only on the element to its right.

The following observations concerning the results exhibited in (17) are pertinent. We have already observed that Γ_z plays the role of a weight operator in the eigenvalue problem (2) defining Φ_m . Therefore, despite the fact that L_F is hermitian, the orthogonality relations for the Φ_m , based on (9), are not the usual relations for an orthogonal eigenfunction set but take the form of "biorthogonality" relations between the eigenfunctions Φ_{α} and the "adjoint" eigenfunctions $\Gamma_z \Phi_{\beta}^*$. This biorthogonal character of the orthogonality relations is also reflected in (17) where each of the representations for I_t is seen to be in the form of a sum of projection operators (i.e., operators with the property $A^2 = A$). A further discussion of this interpretation of the completeness relations (17) and the identification of the eigenmanifolds associated with the projection operators is given elsewhere.¹⁷

THE GREEN'S FUNCTIONS

The steady-state Maxwell equations for the transverse fields excited by a distribution of transverse current sources are here written^{13,16}

$$\left(L - \frac{1}{i} \frac{\partial}{\partial z} \Gamma_z \right) \xi(z) = -i \mathbf{n}(z), \quad (18)$$

where

$$\xi(z) \rightarrow \begin{pmatrix} E_t(z) \\ i H_t(z) \end{pmatrix} \quad \mathbf{n}(z) \rightarrow \begin{pmatrix} J_t(z) \\ i M_t(z) \end{pmatrix} \quad (19)$$

and $J_t(z)$ and $M_t(z)$ represent the distributions of transverse electric and magnetic current sources, respectively. The modification in $\mathbf{n}(z)$ required when there are longitudinal current sources present is given else-

where.^{13,16} To effect the reduction of the inhomogeneous problem (18) to an equivalent point source excitation problem we introduce an operator Green's function $\mathcal{G}(z, z')$ via the requirement

$$\xi(z) = -i \int_{-\infty}^{\infty} \mathcal{G}(z, z') \cdot \mathbf{n}(z') dz'. \quad (20)$$

Assuming the validity of the required interchange of differentiation and integration operations, etc., it is evident that substitution from (20) into (18) yields

$$\left(L - \frac{1}{i} \frac{\partial}{\partial z} \Gamma_z\right) \mathcal{G}(z, z') = I_t \delta(z - z') \quad (21)$$

as the equation which $\mathcal{G}(z, z')$ must satisfy. In this equation, $\delta(z - z')$ is the unit impulse function. For an infinite uniform waveguide the z dependence in (21) may be eliminated by the introduction of Fourier integral representations for all z -dependent quantities. In particular, we represent $\mathcal{G}(z, z')$ as follows:

$$G(z, z') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathcal{G}(\kappa) e^{i\kappa(z-z')} d\kappa \quad (\text{Im } \kappa = 0). \quad (22)$$

Again assuming the validity of the required interchange of differentiation and integration operations, we substitute from (22) into (21) to obtain

$$(L - \kappa \Gamma_z) \mathcal{G}(\kappa) = I_t \quad (23)$$

as the equation which $\mathcal{G}(\kappa)$ must satisfy. This equation does not provide a unique specification for $\mathcal{G}(\kappa)$. It has been shown¹⁶ that a unique specification for $\mathcal{G}(\kappa)$ is obtained only when we require that its adjoint be a solution of the adjoint problem:

$$(L - \kappa^* \Gamma_z) \mathcal{G}^+(\kappa) = I_t. \quad (24)$$

This statement of the adjoint problem exploits the fact that we are here dealing exclusively with hermitian operators $L = L^+$.

By combining the information contained in (2), (13) and (17) with the defining equations (23) and (24) it is readily verified that the following constitute the modal representations for $\mathcal{G}_F(\kappa)$ and $\mathcal{G}_F^+(\kappa)$ in terms of the modes Φ_m of the dissipationless anisotropic waveguide (i.e., in terms of the eigenfunctions of $L_F = L_F^+$):

$$\begin{aligned} \mathcal{G}_F(\kappa) &= - \sum_{\alpha} \frac{1}{M_{\alpha}(\kappa - \kappa_{F,\alpha})} \Phi_{\alpha} \Phi_{\alpha}^* \\ \mathcal{G}_F^+(\kappa) &= - \sum_{\alpha} \frac{1}{M_{\alpha}^*(\kappa - \kappa_{F,\alpha})^*} \Phi_{\alpha}^* \Phi_{\alpha} \end{aligned} \quad (25)$$

where the summation convention here is identical with that for (13).

In an earlier paper,¹⁵ the author has shown that the path of integration specified in the transform relation (22) can be deformed so that the result of the integration is to yield $\mathcal{G}(z, z')$ in terms of two residue series, one

valid for $z < z'$, the other for $z > z'$. For $\mathcal{G}_F(z, z')$ the two residue series are:²¹

$$\begin{aligned} \mathcal{G}_{F>}(z, z') &= -i \left\{ \frac{1}{M_1} \Phi_1 \Phi_1 e^{i\kappa_{F,1}(z-z')} \right. \\ &\quad \left. + \sum'_m \frac{1}{M_m} \Phi_m \Phi_m^* e^{i\kappa_{F,m}(z-z')} \right\} \quad (z > z') \\ \mathcal{G}_{F<}(z, z') &= i \left\{ \frac{1}{M_{-1}} \Phi_{-1} \Phi_{-1} e^{i\kappa_{F,-1}(z-z')} \right. \\ &\quad \left. + \sum'_m \frac{1}{M_m^*} \Phi_m^* \Phi_m e^{i\kappa_{F,m}^*(z-z')} \right\} \quad (z < z') \end{aligned} \quad (26)$$

for the special case which we are here considering, i.e., where $\kappa_{F,\pm 1}$ real and $\Phi_{\pm 1}$ such that $M_1 > 0$ and $M_{-1} < 0$ and where it is understood that $\text{Im } \kappa_{F,m} > 0$ for $m \neq \pm 1$ so that $\text{Im } \kappa_{F,m}^* = \text{Im } \kappa_{F,m} < 0$. The prime used in connection with the summation symbols in (26) indicates that these sums require contributions only from those $\kappa_{F,m}$ for which $\text{Im } \kappa_{F,m} > 0$. A corresponding result is readily obtained for the Green's function $\mathcal{G}_A(z, z')$ of the homogeneous isotropic waveguide. By exploiting the special properties of the modes and modal propagation constants of a homogeneous isotropic waveguide, we can write this corresponding result:

$$\begin{aligned} \mathcal{G}_A(z, z') &= -i \left\{ \frac{1}{N_1} \Psi_{\pm 1} \Psi_{\pm 1} e^{\pm i\kappa_{A,1}(z-z')} \right. \\ &\quad \left. + \sum'_n \frac{1}{N_n} \Psi_{\pm n} \Psi_{\mp n} e^{\pm i\kappa_{A,n}(z-z')} \right\} \quad (z \gtrless z'). \end{aligned} \quad (27)$$

Here we have again separated the contributions from the propagating and nonpropagating modes. The result in (27) is consistent with the conventions adopted earlier in the discussion preceding (7).

THE INTEGRAL EQUATION; THE SCATTERING COEFFICIENTS

We are now in possession of all the tools we require to obtain the integral equation for the discontinuity at the junction plane (or, as we now prefer to designate it, the aperture plane) in Fig. 1. The integral equation is obtained by exploiting the requirement that the transverse fields be continuous across the aperture plane. Suppose we designate the transverse fields in the aperture plane as $\xi(0)$, where $\xi(z)$ is defined in (19). Since

$$\Gamma_z \xi(0) \rightarrow \begin{pmatrix} H_t(0) \times z_0 \\ iz_0 \times E_t(0) \end{pmatrix}, \quad (28)$$

²¹ The Green's operator in this paper is the transverse-transverse part of that in Bresler, *op. cit.*, footnote 15. Since the transverse field components are the independent field components, the Green's operator of this paper is the independent part of that in the earlier work and, therefore, the conclusions arrived at previously are equally valid here.

it follows from an examination of Fig. 2 that $\Gamma_s \xi(0)$ plays the role of an induced-current element at a combined electric and magnetic wall (*i.e.*, a surface on which the tangential components of both the electric and magnetic fields are required to vanish). Thus, if we introduce such a combined electric and magnetic wall into the aperture plane and choose the induced currents in the manner indicated in Fig. 3, these will exhibit the required continuity of the transverse fields across the aperture plane. (The difference in sign arises from the fact that z_0 is an outward normal on the 0_- side of the aperture plane, an inward normal on the 0_+ side.) The integral equation is now obtained by employing (20) to write the total field in each waveguide as the incident field plus the fields excited by the equivalent sources in the aperture plane and then imposing the continuity requirement for the total fields. This yields as the desired integral equation:

$$a_1 \Psi_1 - a_2 \Phi_{-1} = G_0 \Gamma_s \xi(0), \quad (29)$$

where

$$\begin{aligned} \mathcal{G}_0 &= i[\mathcal{G}_{<A}(0, 0) + \mathcal{G}_{>F}(0, 0)] \\ &= \frac{1}{N_1} \Psi_{-1} \Psi_{-1} + \frac{1}{M_1} \Phi_1 \Phi_1 + \sum_n' \frac{1}{N_n} \Psi_{-n} \Psi_n \\ &\quad + \sum_m' \frac{1}{M_m} \Phi_m \Phi_m^*. \end{aligned} \quad (30)$$

For the purposes of the discussion which follows it is convenient to adopt a special notation to distinguish the incident and reflected propagating modes of the two waveguides. Thus, we identify the incident modes as

$$e_1 \equiv \Psi_1 \quad e_2 \equiv \Phi_{-1} \quad (31)$$

and the reflected (or scattered) modes as

$$f_1 \equiv \Psi_{-1} \quad f_2 \equiv \Phi_1. \quad (32)$$

The integral equation (29) will now be replaced by a pair of integral equations for the partial fields associated

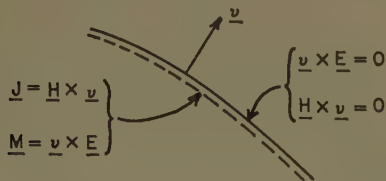


Fig. 2—Induced currents at a combined electric and magnetic wall.

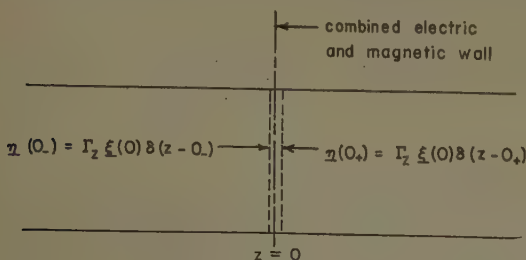


Fig. 3—Induced currents in the aperture plane.

with the two independent excitations. This is accomplished by defining the partial fields $g_{1,2}$ via

$$\xi(0) = a_1 g_1 - a_2 g_2. \quad (33)$$

When we substitute this definition into (29) and exploit the independence of the excitation coefficients $a_{1,2}$ we obtain the following pair of integral equations for the determination of the partial fields:

$$e_i = \mathcal{G}_0 \Gamma_s g_i \quad i = 1, 2. \quad (34)$$

No attempt will be made to obtain explicit solutions to these integral equations. Instead, they will serve to determine variational expressions for the scattering coefficients which describe the effect of the junction on the propagating modes of the two waveguides. The scattering matrix R is introduced via

$$B = RA; \quad R \equiv (R_{ij}) \quad i, j = 1, 2, \quad (35)$$

where

$$A = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad B = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \quad (36)$$

and the coefficients a_i and b_i are defined in Fig. 1. Since all these amplitude coefficients are referred to $z=0$, we can write that

$$\begin{aligned} b_1 &= \frac{1}{N_1} (\Psi_{-1}, \Gamma_s \xi(0)) = -\frac{1}{N_1} (f_1, \Gamma_s \xi(0)) \\ b_2 &= \frac{1}{M_1} (\Phi_1, \Gamma_s \xi(0)) = \frac{1}{M_1} (f_2, \Gamma_s \xi(0)). \end{aligned} \quad (37)$$

By combining the information in (33), (35) and (37), we obtain the following expression for the "normalized" scattering coefficients r_{ij} :

$$r_{ij} = (f_i, \Gamma_s g_j) \quad i, j = 1, 2, \quad (38)$$

where

$$(r_{ij}) = \begin{pmatrix} -N_1 & 0 \\ 0 & M_1 \end{pmatrix} \cdot (R_{ij}) \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (39)$$

VARIATIONAL EXPRESSIONS FOR THE SCATTERING COEFFICIENTS

The integral equations (34) call for the determination of the unknown aperture partial fields g_i (or $\Gamma_s g_i$) with the e_i in the role of the known functions. It is evident from (38) that the scattering coefficients take the form of integral measures (with respect to the known functions f_i) of the unknown functions g_i . The integral equations (34) can, therefore, be employed as the basis for the formulation of variational expressions for the determination of the scattering coefficients. To obtain these variational expressions we introduce the auxiliary integral equations

$$f_i = \mathcal{G}_0^+ \Gamma_s h_i \quad i = 1, 2, \quad (40)$$

in which the f_i now play the role of the known functions and we have introduced a second pair of unknown func-

tions, \mathbf{h}_i . Note that $\phi_0^+ \neq \phi_0^-$. Let

$$\phi = \Gamma_z \phi_0 \Gamma_z. \quad (41)$$

It follows from (34), (38) and (40) that

$$\mathbf{r}_{ij} = (\mathbf{f}_i, \Gamma_z \mathbf{g}_j) = (\mathbf{h}_i, G \mathbf{g}_j) = (\mathbf{h}_i, \Gamma_z \mathbf{e}_j). \quad (42)$$

Thus, we may write the following expression for \mathbf{r}_{ij} :

$$\frac{1}{\mathbf{r}_{ij}} = \frac{(\mathbf{h}_i, G \mathbf{g}_j)}{(\mathbf{h}_i, \Gamma_z \mathbf{e}_j)(\Gamma_z \mathbf{f}_i, \mathbf{g}_j)}. \quad (43)$$

It is readily verified that this expression is variational, *i.e.*, that it is stationary for first-order variations of \mathbf{g}_j and \mathbf{h}_i about the correct solutions of (34) and (40). To employ this variational expression we are required to choose two sets of trial functions. These trial functions are not completely unrelated since (42) can be interpreted as a set of constraints which the trial functions should satisfy. An examination of (34), (40) and (42) makes evident that if, *e.g.*, \mathbf{g}_i is chosen as $\mathbf{e}_k + \alpha \mathbf{f}_i$, then the proper \mathbf{h}_i to be associated with this \mathbf{g}_i is $\mathbf{f}_k + \beta \mathbf{e}_i$, etc. We note also from (42) or (43) that if a self-consistent set of trial functions is desired, then these must be chosen so that no \mathbf{g}_i is orthogonal to any \mathbf{f}_j and, correspondingly, no \mathbf{h}_i is orthogonal to any \mathbf{e}_j .

The variational expression (43) is subject to one rather serious criticism stemming from the nonhermitian character of the operator G . Since G is not hermitian, it is not appropriate to ask whether it is either positive or negative definite. Thus, we cannot exploit the property, so often available in isotropic waveguide problems, that the variational expression is not only stationary for first-order variations about the correct aperture partial fields, but also assumes its maximum (or minimum) value at the correct aperture partial fields. We are, therefore, left without any "built-in" means for gauging the reliability of the variational expression (43).

EXAMPLE—THE DISCONTINUITY AT THE INPUT TO A TRANSVERSELY MAGNETIZED FERRITE-FILLED RECTANGULAR WAVEGUIDE

The problem which we will now consider is that illustrated in Fig. 1 with the infinite uniform waveguide specialized to a rectangular waveguide with perfectly conducting walls. The dimensions of this waveguide are indicated in Fig. 4. To the left of $z=0$ this waveguide is empty, while to the right of $z=0$ it is completely filled

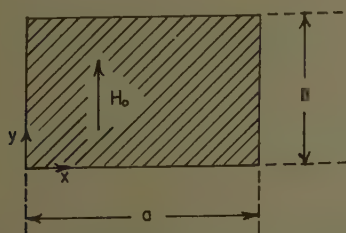


Fig. 4—Transversely magnetized ferrite-filled rectangular waveguide.

with a dissipationless ferrite which is uniformly magnetized in the y direction (see Fig. 4) by an internal dc magnetic field H_0 . It is assumed that the frequency is such that the empty rectangular waveguide propagates only the TE_{10} mode pair (*i.e.*, one such mode in each direction along z). In the numerical example given later, the ferrite parameters are chosen so that the ferrite-filled waveguide likewise propagates only a single pair of modes. The solution to the discontinuity problem will involve only the TE_{n0} modes of both waveguides,¹³ *i.e.*, only those modes which are characterized by the absence of any dependence on y . The nonzero field components of the TE_{n0} modes of both waveguides are E_y , H_x and H_z .

The ferrite will be described by its permittivity $\epsilon\epsilon_0$ and by the effective permeability parameters (defined below) $\nu_1\mu_0$ and $\nu_2\mu_0$, where ϵ_0 and μ_0 are, respectively, the permittivity and permeability of free space. With the dc magnetic field applied along y , the permeability dyadic for the ferrite, $\mu_0\mathbf{y}$, is given by^{7,13}

$$\mathbf{y}_{xyz} \rightarrow \begin{pmatrix} \mu_1 & i\mu_2 & 0 \\ -i\mu_2 & \mu_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (44)$$

where the subscripts associated with \mathbf{y} indicate the cyclic order of the coordinates employed to obtain the indicated representation.²² For our purposes, it turns out to be more convenient to deal with the elements of the inverse \mathbf{y} tensor

$$\mathbf{y}_{xyz}^{-1} \rightarrow \begin{pmatrix} \frac{1}{\nu_1} & -i & \frac{1}{\nu_2} & 0 \\ i & \frac{1}{\nu_2} & \frac{1}{\nu_1} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (45)$$

where ν_1 and ν_2 have already been referred to as the "effective permeability parameters." For a dissipationless ferrite, expressions for μ_1 and μ_2 as functions of

$$\sigma = \frac{\gamma H_0}{\omega}, \quad \rho = \frac{\gamma(4\pi M_s)}{\omega}, \quad (46)$$

where $(4\pi M_s)$ is the saturation magnetization of the ferrite and γ is the magnitude of the gyromagnetic ratio,²³ are available in the literature.⁷ From these we can readily obtain corresponding expressions for ν_1 and ν_2 as follows:

$$\begin{aligned} \nu_1 &= \frac{\mu_1^2 - \mu_2^2}{\mu_1} = \frac{(\rho + \sigma)^2 - 1}{\sigma(\rho + \sigma) - 1} \\ \nu_2 &= \frac{\mu_1^2 - \mu_2^2}{\mu_2} = -\frac{1}{\rho} [(\rho + \sigma)^2 - 1]. \end{aligned} \quad (47)$$

²² This representation differs from that usually found in the literature, *e.g.*, in Tyrus and Held, *op. cit.* footnote 7, because we are using the time dependence $\exp -i\omega t$ instead of $\exp j\omega t$.

²³ $\gamma = 5.6\pi$ when H_0 is expressed in kilo-oersted, $4\pi M_s$ in kilogauss and the frequency in kmc.

Sketches of ν_1 and ν_2 as functions of σ for fixed ρ are shown in Fig. 5. These sketches show the variation of ν_1 and ν_2 as functions of H_0 for a given ferrite material at a fixed frequency. Eqs. (47) apply only for a saturated ferrite and, therefore, any reference to $\sigma=0$ is to be understood to imply σ approaching zero but remaining larger than (the usually small) value corresponding to saturation. Note that the product $\rho\sigma$ is always positive since H_0 and M_s are in the same direction. Therefore, it is sufficient to consider only $\sigma>0$, $\rho>0$ with the understanding that reversing the direction of the dc magnetic field leaves ν_1 unchanged and changes the sign of ν_2 .

The propagation constants of the TE_{n0} modes of the ferrite-filled waveguide are given as:^{7,12,13}

$$\theta_{\pm 1} = \pm \left| \sqrt{\epsilon\nu_1 - \left(\frac{\pi}{K}\right)^2} \right|$$

$$\theta_{\pm m} = \pm i \left| \sqrt{\left(\frac{m\pi}{K}\right)^2 - \epsilon\nu_1} \right| \quad m = 2, 3, 4 \dots, \quad (48)$$

where

$$\theta = \frac{\omega_F}{k} \quad k^2 = \omega^2 \mu_0 \epsilon_0 \quad K = ka. \quad (49)$$

The subscript F will not be exhibited in association with θ since we will not employ a corresponding symbol for the empty waveguide modal propagation constants. The restriction to a single pair of propagating modes implies that $\pi^2 < K^2 \epsilon \nu_1 < 4\pi^2$. Note that the modal propagation constants here occur in positive and negative pairs so that, for the cutoff modes ($m>1$), $\theta_m^* = -\theta_m$ and, therefore, Φ_m^* can now be written as Φ_{-m} . This does not imply that the TE_{n0} modes of the ferrite-filled waveguide exhibit the reflection symmetry property, *i.e.*, we cannot identify Φ_{-m} with $\Gamma_s \Phi_m$. The absence of this property is readily verified by examining the following representation for the Φ_m :

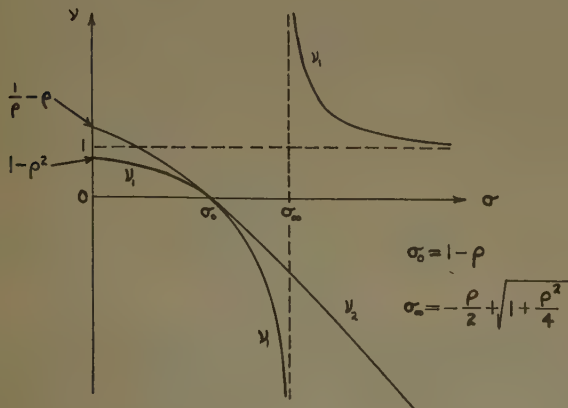


Fig. 5—Ferrite permeability parameters.

$$\Phi_{\pm m} \rightarrow \begin{pmatrix} E_y \\ iH_x \end{pmatrix} = \sqrt{\frac{2}{ab}} \begin{bmatrix} \sin \frac{m\pi x}{a} \\ -i \left[\pm \frac{\theta_m}{|\nu_1|} \sin \frac{m\pi x}{a} - \frac{m\pi}{K\nu_2} \cos \frac{m\pi x}{a} \right] \end{bmatrix}, \quad (50)$$

where $m=1, 2, 3, \dots$. The validity of this representation can be verified from the information concerning the fields of these modes given in the literature.^{7,12,13} In this representation, the magnetic field has been normalized so that the characteristic impedance of free space is unity.²⁴ Only the nonzero transverse field components are exhibited in the representation (50). Corresponding to this representation, that for Γ_s in (3) becomes

$$\Gamma_s \rightarrow \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}. \quad (51)$$

Substitution from (50) and (51) into (12) yields

$$M_{\pm m} = \pm 2 \frac{\theta_m}{|\nu_1|} \quad (52)$$

for the normalization constants of these modes.

Information concerning the TE_{n0} modes of the empty rectangular waveguide can be obtained directly from (48), (50) and (52) by setting $\epsilon=1$, $\nu_1=1$ and $\nu_2=\infty$. In designating the empty waveguide modes, $\Phi_{\pm m}$ becomes $\Psi_{\pm n}$ and $M_{\pm m}$ becomes $N_{\pm n}$. The latter will be designated as

$$N_{\pm n} = \pm 2Y_n, \quad (53)$$

where the Y_n are the characteristic modal admittances.^{13,17}

A study of the representation (50) for the Φ_m reveals that the conjugation identifications exhibited in (7) for the Ψ_n also obtain for this particular set of Φ_m . This conjugation property can be summarized by

$$\Phi_m^* = \Gamma_s \Phi_m^*. \quad (54)$$

It turns out that this property also obtains for the TE_{n0} modes of rectangular waveguides loaded with transversely-magnetized ferrite slabs which do not completely fill the rectangular waveguide. This conjugation property is of interest for the following reasons. In using (43) to evaluate the scattering coefficients, the usual choice for the aperture partial fields will be some superposition of the modes of the two waveguides. The solution for the scattering coefficients will then involve inner products of the form $T_{n,m} = (\Psi_n, \Gamma_s \Phi_m)$. When the indicated conjugation property obtains for both the Ψ_n and Φ_m , it follows that

²⁴ With mks units this means that $H = \sqrt{\mu_0/\epsilon_0} H_{mks}$, where the dimensions of H_{mks} are amperes/meter. The dimensions of H are, therefore, identical with those for E , *i.e.*, volts/meter.

$$T_{n,m}^* = -(\Psi_n^*, \Gamma_s \Phi_m^*) = -(\Psi_n^*, \Gamma_s \Gamma_s \Gamma_s \Phi_m^*) = T_{n,m}^*, \quad (55)$$

since $\Gamma_s^* = -\Gamma_s$, $\Gamma_s^+ = \Gamma_s$, $\Gamma_s \Gamma_s = -\Gamma_s \Gamma_s$, and $\Gamma_s^2 = I_t$. Also, one finds that

$$G^+ = \Gamma_s G^* \Gamma_s, \quad (56)$$

from which it follows that if the subscripts i and j designate propagating modes, then

$$\begin{aligned} (\Phi_i, G\Phi_j) &= (\Phi_j, G^+\Phi_i)^* = (\Phi_j, \Gamma_s G^{*+} \Gamma_s \Phi_i) \\ &= (\Phi_j, G\Phi_i). \end{aligned} \quad (57)$$

In a similar fashion, we can show that $(\Psi_i, G\Psi_j) = (\Psi_j, G\Psi_i)$ and $(\Phi_i, G\Psi_j) = (\Psi_j, G\Phi_i)$. The results in (55) and (57) help to reduce the algebraic labor involved in the use of (43).

We now have all the tools we shall require in the discussion of the discontinuity problem described at the beginning of this section. This problem has been considered previously by Sharpe and Heim¹² who demonstrated that the network in Fig. 6 (when properly interpreted) constitutes the equivalent circuit for the discontinuity. The indicated network consisting of a shunt susceptance iB_s at the junction of two ports, one normalized to the admittance Y_1 , the other normalized to the admittance $\theta_1/|\nu_1|$, is obtained as follows. The empty waveguide is, in the usual fashion, represented by a transmission line of characteristic admittance Y_1 . The voltage and current on this transmission line are directly proportional to the amplitudes of the dominant mode transverse electric and magnetic fields, respectively. In the ferrite-filled waveguide, voltages and currents are defined for the terminal plane at $z=0$ only. The voltage associated with each of the modes $\Phi_{\pm 1}$ is taken as the amplitude of the transverse electric field of the mode, while the current is taken as the amplitude of that portion of the transverse magnetic field which is in space phase with the transverse electric field, since only this component of the transverse magnetic field contributes to the net power flow associated with the mode. Thus, the "characteristic admittance" associated with $\Phi_{\pm 1}$ is $\theta_1/|\nu_1|$ in the sense that this admittance correctly accounts for the power flow associated with these propagating modes. Thus, when the network in Fig. 6 is excited at port 1 only, the ferrite-loaded waveguide behaves like an infinite transmission line of characteristic admittance $\theta_1/|\nu_1|$. Note that this statement makes no implication concerning the transfer relations in the ferrite-loaded waveguide.

The pure shunt character of the discontinuity susceptance indicated in Fig. 6 follows from the fact that, mode for mode, the modal transverse electric fields on both sides of the junction have identical x dependences. This fact was employed by Sharpe and Heim¹² to obtain an integral equation for iB_s , which involves only the electric field in the aperture. In the solution of this integral equation they employed the following approximations:

$$\sqrt{\left(\frac{n\pi}{K}\right)^2 - 1} \approx \sqrt{\left(\frac{n\pi}{K}\right)^2 - \epsilon\nu_1} \approx \frac{n\pi}{K} \quad n = 2, 3, 4 \dots \quad (58)$$

The solution given by Sharpe and Heim was obtained by an iteration about the electric field of the dominant mode. The iteration solution which we shall examine converges when the iteration parameter β (defined below) satisfies $\beta < 1$. The Sharpe and Heim solution for B_s , obtained after a single iteration, is

$$B_s = -\frac{\alpha\beta}{\frac{\pi^2}{4} + \beta} \quad \beta < 1, \quad (59)$$

where

$$\alpha = \frac{\pi}{K} \frac{1 + \nu_1}{\nu_1} \quad \beta = \left(\frac{1}{\nu_2}\right)^2 \left(\frac{\nu_1}{1 + \nu_1}\right)^2. \quad (60)$$

Since the solution given here was obtained as the result of only a single iteration it follows that it will be most accurate when $\beta \ll 1$. When this is the case, we have

$$B_s \approx -\frac{4}{K\pi\nu_2^2} \left(\frac{\nu_1}{1 + \nu_1}\right) \quad \beta \ll 1. \quad (61)$$

The parameter β is shown as a function of σ in Fig. 7. From this sketch it is evident that $\beta < 1$ for $0 < \sigma < 1 - \rho$ and that $\beta \ll 1$ near $\sigma = 0$ provided ρ is sufficiently small. In the following, we will restrict our attention to the indicated range in σ . This is necessary because we re-

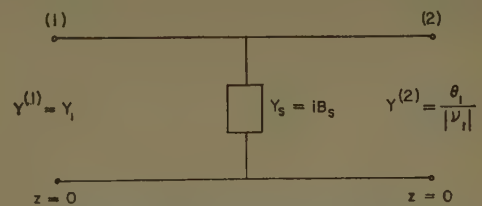


Fig. 6—Equivalent circuit for the junction.

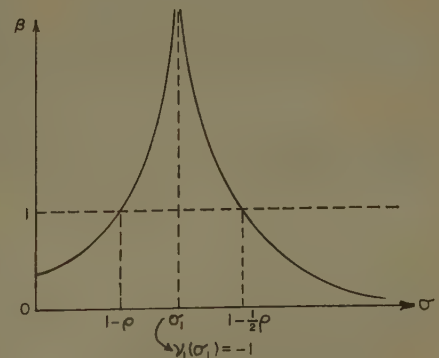


Fig. 7—The Sharpe and Heim parameter β .

quire that ν_1 be positive and, at the same time, sufficiently small so that the ferrite-loaded waveguide will propagate only the lowest mode. Since ν_1 is now restricted to positive values, we may replace $|\nu_1|$ by ν_1 .

To facilitate the comparison of our results, based on the variational expression (43), with those of Sharpe and Heim, we note that the scattering matrix for the junction in Fig. 6 is

$$R = \frac{1}{Y_1 + \frac{\theta_1}{\nu_1} + iB_s} \begin{pmatrix} Y_1 - \frac{\theta_1}{\nu_1} - iB_s & 2\frac{\theta_1}{\nu_1} \\ 2Y_1 & \frac{\theta_1}{\nu_1} - Y_1 - iB_s \end{pmatrix}. \quad (62)$$

From this result it follows that the input admittance \vec{Y}_{in} and \overleftarrow{Y}_{in} seen at ports 1 and 2, respectively, are given in terms of the scattering coefficients by

$$\begin{aligned} \vec{Y}_{in} &= \vec{G}_{in} + i\vec{B}_{in} = \frac{\theta_1}{\nu_1} + iB_s = Y_1 \frac{1 - R_{11}}{1 + R_{11}} \\ \overleftarrow{Y}_{in} &= \overleftarrow{G}_{in} + i\overleftarrow{B}_{in} = Y_1 + iB_s = \frac{\theta_1}{\nu_1} \frac{1 - R_{22}}{1 + R_{22}}. \end{aligned} \quad (63)$$

Since the conductances $\vec{G}_{in} = \theta_1/\nu_1$ and $\overleftarrow{G}_{in} = Y_1$ are known exactly, the procedure which we will adopt in comparing our results with those of Sharpe and Heim is as follows. The scattering coefficients \vec{R}_{11} and \overleftarrow{R}_{22} will be computed via (43). The admittances \vec{Y}_{in} and \overleftarrow{Y}_{in} will then be computed via (63). The computed conductances will then be compared with the indicated exact values and the computed susceptances will be compared with the value of B_s given by (59).

The solutions for R_{11} and R_{22} which we will exhibit below are obtained with the following simple choices of trial functions for the aperture partial fields:

$$\begin{aligned} \text{for } R_{11}: \quad \mathbf{g}_1 &= \Phi_1 & \mathbf{h}_1 &= \Phi_{-1} \\ \text{for } R_{22}: \quad \mathbf{g}_2 &= \Psi_{-1} & \mathbf{h}_2 &= \Psi_1. \end{aligned} \quad (64)$$

Other choices for the trial functions were considered elsewhere.¹³ These other choices included trial functions consisting of various superpositions of the propagating modes of the two waveguides with the relative amplitudes of the modes determined from the requirement that (43) be stationary. The results obtained with these more complicated trial functions do not differ in their essential features from those which we will exhibit below.

Substitution of the trial functions in (64) into the variational expressions (43) yields, via (39), the following results for R_{11} and R_{22} :

$$\begin{aligned} R_{11} &= -\frac{r_{11}}{N_1} = -\frac{T_{1,-1}T_{-1,1}}{N_1(\Phi_{-1}, G\Phi_1)} \\ R_{22} &= -\frac{r_{22}}{M_1} = -\frac{T_{1,-1}T_{-1,1}}{M_1(\Psi_1, G\Psi_{-1})}, \end{aligned} \quad (65)$$

where

$$\begin{aligned} (\Phi_{-1}, G\Phi_1) &= \frac{T_{-1,1}T_{-1,-1}}{N_1} + \sum_{n=2}^{\infty} \frac{T_{n,1}T_{n,-1}}{N_n} \\ (\Psi_1, G\Psi_{-1}) &= \frac{T_{1,1}T_{-1,1}}{M_1} + \sum_{m=2}^{\infty} \frac{T_{1,m}T_{-1,m}}{M_m}. \end{aligned} \quad (66)$$

Introducing the representations for the Ψ_n and Φ_m into the definition of $T_{n,m}$, and performing the required integrations, yields

$$\begin{aligned} T_{\pm 1,1} &= \pm Y_1 + \frac{\theta_1}{\nu_1} & T_{1,\pm 1} &= Y_1 \pm \frac{\theta_1}{\nu_1} \\ T_{n,\pm 1} &= -\frac{2}{K\nu_2} [1 + (-1)^n] \frac{n}{n^2 - 1} & (n > 1) \\ T_{\pm 1,m} &= \frac{2}{K\nu_2} [1 + (-1)^m] \frac{m}{m^2 - 1} & (m > 1). \end{aligned} \quad (67)$$

Note that $T_{n,\pm 1}$ is zero when $n > 1$ is odd and, similarly, $T_{\pm 1,m}$ is zero when $m > 1$ is odd. Note also that $T_{n,1} = T_{n,-1}$ and $T_{1,m} = T_{-1,m}$. Thus, the two infinite series involved in (66) become

$$\begin{aligned} \sum_{n=2}^{\infty} \frac{T_{n,1}T_{n,-1}}{N_n} &= -\frac{i}{2} S_n \\ S_n &= \frac{2}{K\pi\nu_2^2} \sum_{n=1}^{\infty} \frac{n^2}{\left(n^2 - \frac{1}{4}\right)^2 \sqrt{n^2 - \left(\frac{K}{2\pi}\right)^2}} \\ \sum_{m=2}^{\infty} \frac{T_{1,m}T_{-1,m}}{M_m} &= -i \frac{\nu_1}{2} S_m \\ S_m &= \frac{2}{K\pi\nu_2^2} \sum_{m=1}^{\infty} \frac{m^2}{\left(m^2 - \frac{1}{4}\right)^2 \sqrt{m^2 - \epsilon\nu_1 \left(\frac{K}{2\pi}\right)^2}}. \end{aligned} \quad (68)$$

If these series are summed by employing the Sharpe and Heim approximation given in (58) the result is

$$S_n \approx S_m \approx S_0 = \frac{2}{K\pi\nu_2^2} \sum_{n=1}^{\infty} \left(\frac{n}{n^2 - \frac{1}{4}} \right)^2 = \frac{4}{K\pi\nu_2^2}. \quad (69)$$

The two series S_n and S_m can be summed as accurately as desired, by the addition of successive corrections to S_0 . Thus, for example, for S_n we write

$$\begin{aligned} S_n &= \frac{2}{K\pi\nu_2^2} \left\{ 2 + \frac{16}{9} \left| \frac{1}{\sqrt{1 - \left(\frac{K}{2\pi}\right)^2}} - 1 \right| \right. \\ &\quad \left. + \frac{32}{225} \left| \frac{1}{\sqrt{1 - \left(\frac{K}{4\pi}\right)^2}} - 1 \right| + \dots \right\}. \end{aligned} \quad (70)$$

For S_m we write a similar expression with $(K/2n\pi)^2$ in the radicals replaced by $\epsilon\nu_1(K/2m\pi)^2$. For the numerical example which we will consider below, S_m was summed with the two corrections indicated above while a single

correction was found to be adequate for S_n .

Introduction of the information developed above into (65) yields the following results for R_{11} and R_{22} :

$$R_{11} = \frac{\left(Y_1 - \frac{\theta_1}{\nu_1}\right)^2}{Y_1^2 - \left(\frac{\theta_1}{\nu_1}\right)^2 - iY_1 S_n}$$

$$R_{22} = \frac{\left(Y_1 - \frac{\theta_1}{\nu_1}\right)^2}{Y_1^2 - \left(\frac{\theta_1}{\nu_1}\right)^2 + i\theta_1 S_m} \quad (71)$$

The admittances \vec{Y}_{in} and \overleftarrow{Y}_{in} corresponding to these results are found to be

$$\vec{G}_{in} = \frac{\theta_1}{\nu_1} + \left(Y_1 - \frac{\theta_1}{\nu_1}\right) \Delta_n \quad \overleftarrow{G}_{in} = Y_1 - \left(Y_1 - \frac{\theta_1}{\nu_1}\right) \Delta_m$$

$$\vec{B}_{in} = -\frac{1}{2} S_n (1 - \Delta_n) \quad \overleftarrow{B}_{in} = -\frac{\nu_1}{2} S_m (1 - \Delta_m)$$

$$\Delta_n = \frac{S_n^2}{4 \left(Y_1 - \frac{\theta_1}{\nu_1}\right)^2 + S_n^2}$$

$$\Delta_m = \frac{(\nu_1 S_m)^2}{4 \left(Y_1 - \frac{\theta_1}{\nu_1}\right)^2 + (\nu_1 S_m)^2} \quad (72)$$

Numerical results based on these formulas are shown in Figs. 8 and 9. These results were obtained with $\rho=0.4$, $\epsilon=3.0$ and $K=1.25\pi$. The values for ρ and ϵ are both somewhat on the low side compared with the values appropriate for available ferrites. These values were chosen to yield a reasonably large range of values for σ in which the ferrite-loaded waveguide propagates only the dominant mode pair. In Fig. 8, the conductances \vec{G}_{in} and \overleftarrow{G}_{in} computed from (72) are compared with the exact values $\vec{G}_{in}=\theta_1/\nu_1$ and $\overleftarrow{G}_{in}=Y_1=0.6$. This comparison makes evident that the variational solutions are reasonably accurate except in the neighborhood of the value of σ at which the ferrite-filled waveguide dominant mode cuts off (i.e., $\sigma \approx 0.55$).

The results given in Fig. 9, while somewhat disappointing, are not as bad as they appear. In support of this comment we offer the following observations. The susceptance values are generally much smaller than the conductances, so that a large error in B_s does not necessarily imply an equally large error in the scattering coefficient. Since the conductance results were quite good, the implication is that the accuracy of the variational solution is expended almost entirely on the determination of the conductances. The nature of the dependence on σ of the variational solutions is distinctly different from that of the Sharpe and Heim solution in the neighborhood of $\sigma=0.52$. This difference does not

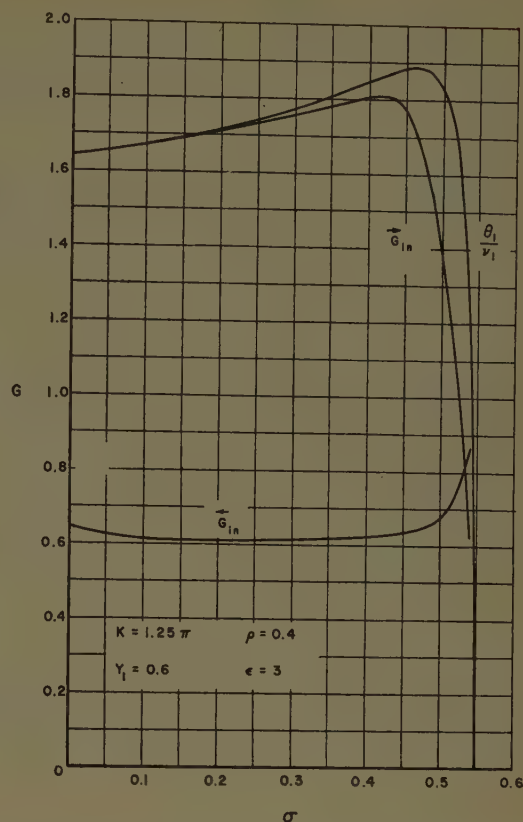


Fig. 8—The variational solutions for \vec{G}_{in} and \overleftarrow{G}_{in} compared with the known correct values $\vec{G}_{in}=\theta_1/\nu_1$ and $\overleftarrow{G}_{in}=Y_1$.

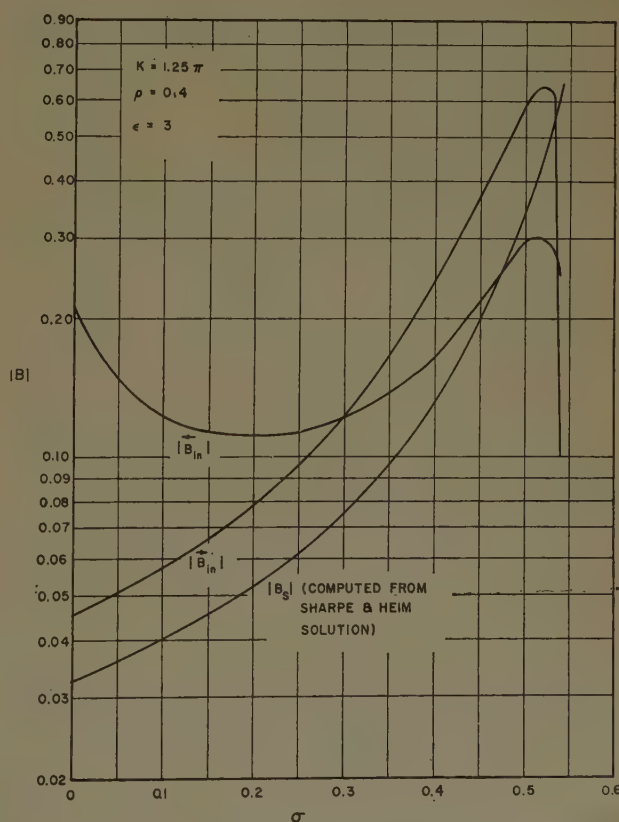


Fig. 9—Capacitive discontinuity susceptance.

cause us any concern since, with $\beta \rightarrow 1$ as $\sigma \rightarrow 0.6$; the latter solution will be least accurate for values of σ near $\rightarrow 0.6$. For $0 < \sigma < 0.5$, the dependence on σ of the \vec{B}_{in} solution is quite similar to that of the Sharpe and Heim solution. If the expression for \vec{B}_{in} in (72) were evaluated by employing the approximation (58), *i.e.*, by replacing S_n by S_0 , the correlation of these two curves would be improved. This result would obtain because S_n and S_0 , in their dependence on σ , differ only by a scale factor since both are inversely proportional to ν_2^2 .

The most disturbing feature of the results shown in Fig. 9 is the fact that the values of susceptance computed from R_{11} and R_{22} differ considerably over most of the range of σ in which computations were made. This difference results from the fact that the dependence of S_m on σ is quite different from that of S_n , since the summands of S_m are functions of ν_1 . This dependence on ν_1 causes S_m to become infinite as ν_1 approaches the value at which the lowest cutoff mode of the ferrite-loaded waveguide would become a propagating mode. We observe that the trial functions we employed to compute R_{11} and R_{22} were such that the cutoff modes of the empty waveguide contributed to the solution for R_{11} only, whereas the cutoff modes of the ferrite-loaded wave-

guide contributed to the solution for R_{22} only. The fact that the Sharpe and Heim solution resembles \vec{B}_{in} more closely than \vec{B}_{in} does not imply that we should favor the R_{11} solution over the R_{22} solution. We recall that the Sharpe and Heim solution embodied the approximation (58). If this approximation were to be used in the evaluation of \vec{B}_{in} , the result would be very similar to \vec{B}_{in} . Thus, at present, we are unable to choose between the solutions \vec{B}_{in} and \vec{B}_{in} . The inability to choose between these two solutions bears out the criticism of the variational solution which was made in the final paragraph of the preceding section.

ACKNOWLEDGMENT

This paper is based on a part of the author's dissertation¹⁸ submitted to the Polytechnic Institute of Brooklyn in partial fulfillment of the requirements for the degree of Doctor of Electrical Engineering. The author wishes to express his gratitude to his thesis adviser, Professor N. Marcuvitz, for his guidance and assistance. Also, the author is indebted to his former colleagues at the Microwave Research Institute for their numerous helpful discussions and criticisms.

A Ferrite-Filled Cylindrical Cavity*

T. S. CHU† AND R. G. KOUYOUMJIAN†

Summary—The modes in a ferrite-filled cylindrical waveguide have been analyzed, and a characteristic equation given for their propagation constants in terms of the frequency and applied magnetic field.¹⁻⁴ However, there have been no reported experimental confirmations of these solutions. It is convenient to carry out the experiment by measuring the related resonances in a ferrite-filled cylindrical cavity. This has been done for the F_{0n0} and F_{0n1} modes in a symmetrically excited cavity. The measured resonant frequencies are in general agreement with numerical results based on the solution of the characteristic equation and the theories of Polder and Rado. Assuming that $\mu_z \approx \mu$, the components of the tensor permeability below the magnetic saturation of the ferrite may be determined over a range of frequencies from the measured curves of resonance. This type of measurement avoids certain difficulties inherent with the perturbation method. The curves of resonance may also be used to calculate the gyromagnetic ratio of the ferrite.

I. INTRODUCTION

RELATIVELY few boundary value problems involving gyromagnetic media can be solved rigorously to the point where numerical results are practicable. Consequently, one is limited in the use of the permittivity and tensor permeability to check measurements involving gyromagnetic media, or conversely, in the design of experiments from which these parameters can be determined. The microwave propagation through a ferrite-filled cylindrical waveguide, however, has been treated by a number of authors, who derived the characteristic equation for the propagation constant. In this paper the solution of the characteristic equation is employed to relate the resonant variables (frequency and magnetic field) for certain modes in a ferrite-filled cylindrical cavity. Expressions for the components of the tensor permeability derived from the theories of Polder and Rado are used in the solution, so that a test of these theories is effected by comparing the calculated resonances with the experimental resonances.

Conversely, the calculated and experimental resonances may be utilized to find the components of the tensor permeability. This type of measurement accomplishes a natural averaging of the ferrite inhomogeneities and thus provides a desirable check on the values measured by the cavity perturbation method, where small samples of the ferrite are used. Furthermore, in contrast

with the cavity perturbation method, this type of measurement can be carried out over a range of frequencies with a single cavity.

II. THEORY

The formal solution of electromagnetic wave propagation in a circular waveguide completely filled with a magnetized ferrite has been presented by Suhl and Walker,¹ Kales,² van Trier,³ and Epstein.⁴ Starting with Maxwell's equations a pair of coupled equations which must be satisfied by the longitudinal components of the microwave field in the waveguide is obtained.

$$\nabla_t^2 E_z + aE_z + bH_z = 0, \quad (1)$$

$$\nabla_t^2 H_z + cH_z + dE_z = 0, \quad (2)$$

where the dependency on time and on the direction of propagation along the z axis is of the form $e^{j(\omega t - \beta z)}$, and the subscript t denotes differentiation with respect to the transverse coordinates. The coefficients a , b , c , d involve the angular frequency ω , the propagation constant β , the permittivity ϵ , and the components of the tensor permeability which for a ferrite magnetized in the z direction have the form

$$\bar{\mu} = \begin{pmatrix} \mu & -j\kappa & 0 \\ j\kappa & \mu & 0 \\ 0 & 0 & \mu_z \end{pmatrix}. \quad (3)$$

Setting

$$\rho_H = \frac{\kappa}{\mu}, \quad \nu_H = \frac{\mu}{\mu_z}, \quad \mu_{\text{eff}} = \frac{\mu^2 - \kappa^2}{\mu}$$

$$a = \omega^2 \epsilon \mu_{\text{eff}} - \beta^2, \quad (4a)$$

$$b = j\beta \omega \mu_z \rho_H, \quad (4b)$$

$$c = \omega^2 \epsilon \mu_z - \beta^2 / \nu_H, \quad (4c)$$

$$d = -j\beta \omega \epsilon \rho_H. \quad (4d)$$

Next letting

$$E_z = \phi_1 + \phi_2, \quad (5a)$$

$$H_z = g_1 \phi_1 + g_2 \phi_2, \quad (5b)$$

and imposing the condition that $g_1 \neq g_2$, (1) and (2) may be transformed into the uncoupled pair

$$\nabla_t^2 \phi_{1,2} + \sigma_{1,2} \phi_{1,2} = 0 \quad (6)$$

where the $\sigma_{1,2}$ are in turn functions of a , b , c , d . All field components can be obtained from $\phi_{1,2}$. After imposing the boundary conditions, a transcendental characteristic equation results from which the propagation constants in a circular waveguide, or the resonant frequencies of a cylindrical cavity, can be obtained in principle.

* The research reported in this paper was sponsored in part by the U. S. Army Signal Res., and Dev. Lab., Fort Monmouth, N. J.
† Antenna Lab., Dept. of Elec. Engrg., Ohio State University, Columbus, Ohio.

¹ H. Suhl and L. R. Walker, "Topics in guided wave propagation through gyromagnetic media I," *Bell Sys. Tech. J.*, vol. 33, pp. 579-659; May, 1954.

² M. L. Kales, "Modes in waveguides that contain ferrites," *J. Appl. Phys.*, vol. 24, pp. 604-608; May, 1953.

³ A. A. T. M. van Trier, "Guided electromagnetic waves in anisotropic media," *Appl. Sci. Res.*, vol. 3, sec. B, no. 5, pp. 305-371; 1953.

⁴ P. S. Epstein, "Theory of wave propagation in a gyromagnetic medium," *Rev. Mod. Phys.*, vol. 28, pp. 3-17; January, 1956.

In order to satisfy the boundary conditions in the circular waveguide, the angular dependence of the solution must be of the form $e^{jm\theta}$, where m is any integer, positive, negative or zero. Suhl and Walker¹ have analyzed the case for $m = \pm 1$. The symmetric modes $m=0$ are investigated here, both theoretically and experimentally, using the resonances in a ferrite-filled cylindrical cavity.

The characteristic equation for the symmetric modes may be written as

$$\frac{F(x_1)}{\lambda_2} = \frac{F(x_2)}{\lambda_2}, \quad (7)$$

where

$$F(x) = \frac{J_1(x)}{xJ_0(x)}. \quad (8)$$

The λ 's satisfy

$$\lambda_{1,2}^2 - \frac{(1 - \nu_H) \left(1 - \frac{\beta^2}{\omega^2 \mu \epsilon}\right) + \nu_H \rho_H^2}{\rho_H} \lambda_{1,2} - \frac{\beta^2}{\omega^2 \mu \epsilon} = 0. \quad (9)$$

The x 's are given by

$$x_{1,2}^2 = \left[\omega^2 \mu \epsilon - \frac{\beta^2}{\nu_H} - \rho_H \omega^2 \mu \epsilon \lambda_{1,2} \right] r_0^2, \quad (10)$$

and r_0 is the radius of the cylindrical cavity.

The symmetric mode resonances can be designated by F_{0np} . The integer n denotes the order of the roots of the characteristic equation and p is the number of half-wavelengths in the axial direction. These modes are not related to the conventional cylindrical cavity modes except when $p=0$; then $\beta=0$, corresponding to the cutoff condition for the F_{0n} modes in a cylindrical waveguide. From (7)–(10) it is seen that with $\lambda_1=0$, $x_2^2 = \omega^2 \epsilon \mu_{\text{eff}} r_0^2$, and the characteristic equation simplifies to

$$J_0(x_2) = 0 \quad (11)$$

for the F_{0n0} modes. Since the zeros of the Bessel functions exist only for real arguments, it is evident that the F_{0n0} resonances occur only for $\mu_{\text{eff}} > 0$ regardless of the cavity dimensions. It can be seen from (4a)–(4d) that (1) and (2) uncouple in the demagnetized state of the ferrite, where $\mu = \mu_z$ and $\kappa = 0$, and thus the hybrid modes of the magnetized ferrite reduce to the conventional TE and TM type modes as one would expect. In the case of the F_{0n1} modes, $p=1$ and $\beta = \pi/h$, where h is the height of the cavity. The graphical solution of (7) for the F_{0n1} modes is discussed in the Appendix.

It is evident that the resonant state of a ferrite-filled cavity must be defined in terms of the applied static magnetic field as well as the frequency. The frequency

and magnetic field determine the values of μ , κ , and μ_z ; moreover, ϵ may be significantly dependent on the frequency. These constitutive parameters in turn determine the values of the resonant frequency as indicated in (7)–(10). For the ranges of frequency and magnetic field to be considered here, the MgMn-ferrite of interest has relatively low loss and so the constitutive parameters are essentially real valued.

In the saturated state of the ferrite the components of the tensor permeability are given by equations based on Polder's theory; when losses can be neglected these equations have the form⁵⁻⁷

$$\frac{\mu}{\mu_0} = \left(1 + \frac{4\pi M_s H^i \gamma^2}{\gamma^2 H^{i2} - \omega^2} \right) \quad (12a)$$

$$\frac{\kappa}{\mu_0} = \frac{4\pi M_s \omega \gamma}{\gamma^2 H^{i2} - \omega^2} \quad (12b)$$

$$\frac{\mu_z}{\mu_0} = 1. \quad (12c)$$

In the above equations M_s is the saturation magnetization, H^i is the static magnetic field within the ferrite, and γ is the gyromagnetic ratio. Thus, in addition to their dependence on ω and H^i , the components of the tensor permeability for a given ferrite in a magnetically saturated state depend upon M_s and γ . It is noted that γ is a negative quantity.

Eqs. (12a)–(12c) are not applicable to the unsaturated state of the ferrite. Rado^{8,9} has proposed a generalization of Polder's theory which will apply to the unsaturated state; this extension leads to the following values for the components of the tensor permeability

$$\mu_s = \mu = \mu_0, \quad (13a)$$

$$\kappa = - \frac{4\pi M \gamma}{\omega} \mu_0, \quad (13b)$$

when the conditions $|\gamma| H^i \ll \omega$ and $4\pi M_s |\gamma| \ll \omega$ are met. In (13b) M is the magnetization related to H^i by the magnetization curve of the ferrite. Since only the first condition is met for the problem under consideration, one might conclude that the above equations cannot be employed here. However, it is evident from van Trier's presentation,³ that the expression for κ should be valid even though the second condition is not met. The as-

⁵ It is customary to use the Gaussian system of units for quantities describing the static magnetization of the ferrite and to use the rationalized MKS system of units for quantities describing the electromagnetic fields.

⁶ D. Polder, "On the theory of ferromagnetic resonance," *Phil. Mag.*, vol. 40, pp. 99–115; January, 1959.

⁷ C. L. Hogan, "Ferromagnetic Faraday effect at microwave frequencies," *Revs. Mod. Phys.*, vol. 25, pp. 253–263; January, 1953.

⁸ G. T. Rado, "Theory of the microwave permeability tensor and Faraday effect in nonsaturated ferromagnetic materials," *Phys. Rev.*, vol. 89, p. 529; January, 1953.

⁹ G. T. Rado, "On the electromagnetic characterization of ferromagnetic media: permeability tensors and spin wave equations," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-4, pp. 512–530; July, 1956.

sumption that $\mu_z \approx \mu$ is plausible because $\mu = \mu_z$ for the demagnetized ferrite, and again μ is known to approximately equal $\mu_z = \mu_0$ when H^i is increased to the point where the ferrite is essentially saturated.¹⁰

In calculating the F_{0n0} and F_{0n1} resonances above magnetic saturation the components of the tensor permeability are found from (12a)–(12c) using measured values of M_s and γ . When the ferrite is below magnetic saturation the value of κ is found from (13b); however, no reliable expression exists for μ . Hence, it was decided to determine μ directly from the F_{0n0} resonances and calculate only the F_{0n1} resonances below magnetic saturation.

III. EXPERIMENT

The resonant frequencies of two cavities completely filled with a MgMn-ferrite (General Ceramics R-1 ferrite) were measured over the X-band frequency range as a function of the static magnetic field applied in the direction of the cavity axis. Both cavities are 3.440 inches in diameter, but their heights differ, one being 0.195 inch and the other 0.160 inch. Each cavity is fed by a small probe 0.01 inch in diameter extending 0.05 inch into the ferrite from the center of one of the end faces. The relationship of the cavity to the complete measuring system is shown in Fig. 1. A Varian X-13

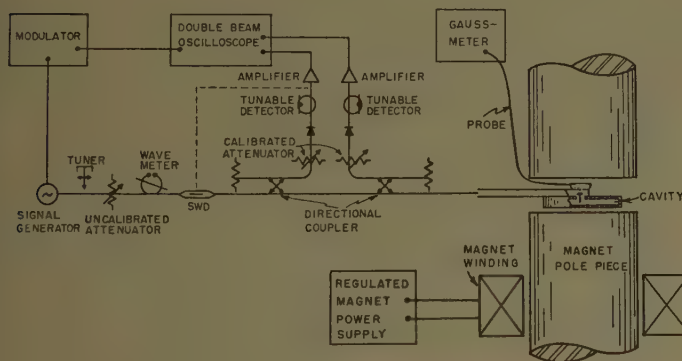


Fig. 1—The measuring system.

klystron is frequency modulated by a 60-cycle sine wave. The cavity resonances appear as pips on the oscilloscope trace. The Q of the cavity was found by a sweep frequency method^{11,12} having a probable error in the measured Q of about 10 per cent. The unloaded Q of the ferrite-filled cavity is about 500 for the well-

matched resonances. This was found to fluctuate irregularly from one resonant point to the next. It was too difficult to measure the Q 's of the barely discernible resonances which were classified as weak resonances. The unloaded Q of the air-filled cavity is about 1700, which compares favorably with the calculated value. The cavity is located in a 1.8 inch gap of an electromagnet with pole faces 5.5 inches in diameter. Rings are located at the edge of the pole faces to improve the uniformity of the magnetic field in the region occupied by the cavity. The magnetic field is measured by a D-79 Dyna-Labs gaussmeter, which was calibrated against a proton resonance gaussmeter. The field variation between the edge and center of the cavity with the ferrite disk in place was found to be no greater than 3 per cent for an applied field of 1000 gauss.

In order to determine the microwave properties of the ferrite, the magnetization curve, gyromagnetic ratio and permittivity were measured. Under conditions which were described in the previous section, measurements of M and γ may be used to determine the components of the tensor permeability, according to the theories of Polder and Rado.

The magnetization curve of the ferrite was measured by a Fahy permeameter using a bar sample and also with a ballistic type of permeameter appropriate for thin rod samples.¹³ An empirical equation for the magnetization similar to that given by Stoner¹⁴ was fitted to the experimental data and the value of $4\pi M_s$ was found to be 2060 gauss.

It is desirable to measure both the gyromagnetic ratio and permittivity as a function of frequency, because a significant decrease in the gyromagnetic ratio for several ferrites as the frequency increases has been reported¹⁵ and because the permittivities of different ferrites have different frequency behavior.¹⁶ The gyromagnetic ratio was measured by putting a number of small ferrite spheres (about 1.5 mm in diameter) in a rectangular waveguide and observing the gyromagnetic resonance absorption. The permittivity was measured by observing the frequency shift of the TE_{101} mode for a rectangular cavity when the field is perturbed by a ferrite sphere 2 mm in diameter located at the center of the cavity. The measured values of both the gyromagnetic ratio and permittivity over X band showed a decrease of a few per cent which is within the estimated experimental error. Therefore, the value of the gyromagnetic ratio $|\gamma|$ is taken to be a constant 18.6×10^6 rad/oersted-second and the value of the relative permittivity ϵ_r is taken to be a constant 12.

¹⁰ R. C. LeCraw, and E. G. Spencer, "Tensor permeabilities of ferrites below magnetic saturation," 1956 IRE CONVENTION RECORD, pt. 5, pp. 66–74.

¹¹ R. C. Ward, "Measurement of Q of Single-Ended Resonators," The Ohio State University Res. Foundation, Wright Air Dev. Center, Wright-Patterson AFB, Dayton, Ohio, Tech. Rept. prepared under Contract W33-038-ac-15162; August, 1953.

¹² E. D. Reed, "A sweep frequency method of Q measurement for single-end resonators," *Proc. Natl. Electronics Conf.*, vol. 7, p. 162; 1951.

¹³ R. Bozorth, "Ferromagnetism," D. Van Nostrand Co., New York, N. Y., p. 843; 1951.

¹⁴ Stoner, "Magnetization curves," *Repts. Prog. Phys.*, vol. 13, pp. 83–183; 1950.

¹⁵ J. Snider, "Ferromagnetic resonance in polycrystalline ferrites," *Appl. Sci. Res.*, vol. 7, sec. B, no. 3, p. 185; 1958.

¹⁶ L. G. Van Uitert, "Dielectric properties of and conductivities in ferrites," *Proc. IRE*, vol. 44, pp. 1294–1303; October, 1956.

IV. DISCUSSION

A. Curves of Resonance

The calculated curves of resonance for the ferrite-filled cavity 0.160 inch thick are shown in Fig. 2. The corresponding measured curves of resonance are shown in Fig. 3. The magnetization of the ferrite is defined in terms of the applied flux density B^i within the ferrite rather than the internal applied magnetic field intensity H^i . The reason for this is that B^i can be more accurately computed from the measurement of the external field than can H^i , and M used in (13b) can be more accurately found from B^i than H^i . When the magnetization of the ferrite is carried out in a prescribed manner, its magnetic state can be defined either by B^i or H^i . It should be noted that B^i is numerically equal to the external applied magnetic field intensity H^e . The range of B^i covered in Figs. 2 and 3 lies well below ferromagnetic resonance ($B_{\text{res}}^i = 5440$ gauss at 10 kmc). In Fig. 2 the calculated limit of the region of the F_{0n0} modes is given by the line $\mu_{\text{eff}} = 0$. The region of ferrite magnetic saturation lies above the line labeled saturation. The point at which the maximum static permeability occurs is indicated by the line labeled static μ_{max} .

The F_{0n0} resonances appear as the family of curves with positive slope. It can be seen from (11) and (12) that above magnetic saturation the resonant angular frequency ω_n may be put into the form

$$\omega_n^2 = \frac{\gamma^2 B^{i^2}}{2} + \frac{\Omega_n^2}{2\epsilon_r} \pm \sqrt{\left(\frac{\gamma^2 B^{i^2}}{2} + \frac{\Omega_n^2}{2\epsilon_r}\right)^2 - \frac{\Omega_n^2}{\epsilon_r} \gamma^2 B^i H^i}, \quad (14)$$

where Ω_n is the resonant angular frequency for an air-filled cavity having the same dimensions as the ferrite-filled cavity, and the $+$ sign applies to the resonances below ferromagnetic resonance. It is evident that as B^i increases the resonant curves asymptotically approach the line $\omega = |\gamma| B^i$. The approach is more rapid for the modes of lower order. This property will be used to check the gyromagnetic ratio found from the waveguide measurement. It can be shown that as μ_{eff} approaches zero, the Q of the cavity decreases rapidly with the result that the lower order resonances are very weak. In Fig. 3 it is seen that the first and second order resonances close to the line $\mu_{\text{eff}} = 0$ are not detected and that the third and fourth order resonances are weak.

The F_{0n1} resonances appear as a family of curves with negative slope as shown in Fig. 2. Above magnetic saturation these curves become much less dependent upon the applied magnetic field. The explanation for this may be found by examining (7) through (10), where $\rho_H = \kappa/\mu$ is the only quantity which varies with B^i ($\nu_H \approx 1$), and it can be shown from (12a) and (12b) that ρ_H changes very

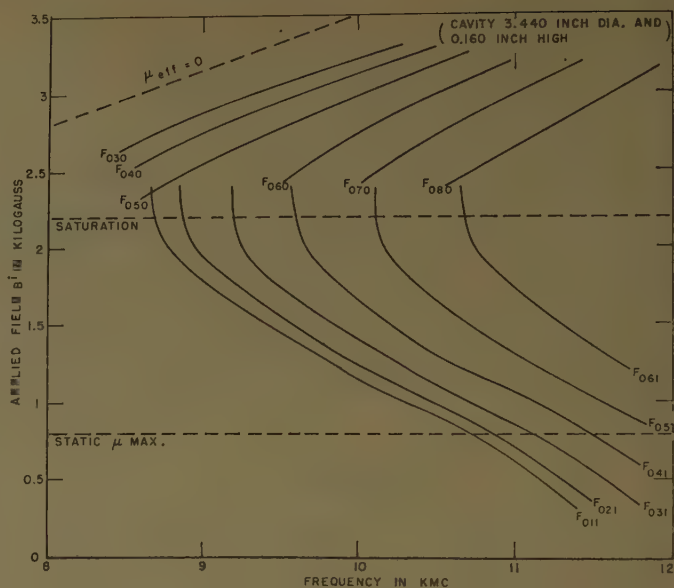


Fig. 2—The calculated curves of resonance for a ferrite-filled cavity.

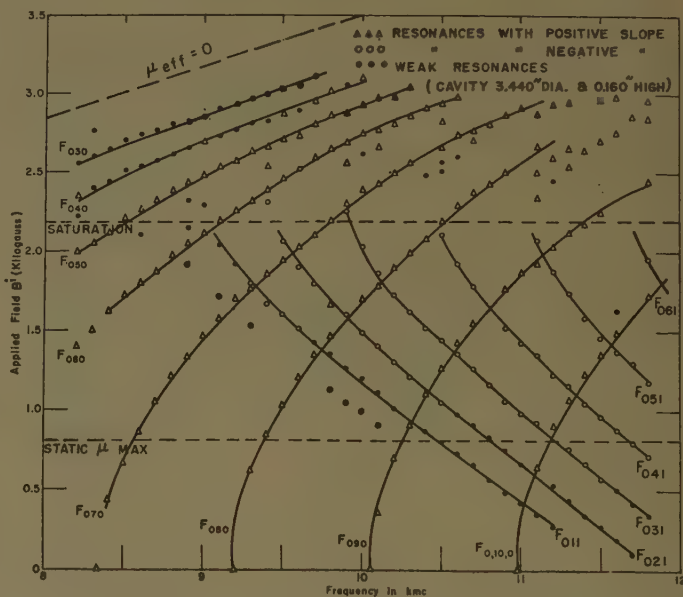


Fig. 3—The measured curves of resonance for a ferrite-filled cavity.

slowly with B^i in the range of interest. Referring to Fig. 3, it is rather curious that the F_{0n1} resonances are not detected above magnetic saturation. Plausible explanations for this are low Q 's above magnetic saturation and the general difficulty of detecting resonances which are of nearly constant frequency when the measurements are carried out at constant frequency.

Comparing the calculated curves of resonance in Fig. 2 with the measured curves in Fig. 3 it is seen that there is quite good agreement between the two for the F_{0n0} modes above magnetic saturation. Thus, (12a)–(12c), derived from Polder's theory, may be used to obtain reasonably accurate results for a problem of this type. The well separated, measured F_{0n0} resonant curves be-

low magnetic saturation greatly facilitate the identification of the order n . On the other hand, qualitative agreement between the two sets of resonant curves for the F_{0n1} modes below magnetic saturation primarily indicates that when (13a) and (13b) derived from Rado's theory, are applied to a problem of this type, less satisfactory results can be obtained.

An interesting point is that by varying the static magnetic field the two types of modes, F_{0n0} and F_{0n1} , can be brought arbitrarily close to each other in terms of frequency, yet remain distinct.

B. Tensor Permeability and Gyromagnetic Ratio

The curves of resonance for the F_{0n0} and F_{0n1} modes are two independent functions of the tensor permeability components. Assuming that $\mu = \mu_z$, the two families of curves in Fig. 3 can be employed to determine μ and κ as functions of both the magnetic field and frequency. The numerical calculation was carried out by a graphical method similar to that described in the Appendix. The quantities κ/μ and $\epsilon\mu$ were found first; then using the measured value of ϵ , κ and μ may be plotted as functions of B^i and f , as shown in Figs. 4 and 5. Fig. 4 indicates fair agreement between the value of κ found from (13b) and that found from the cavity resonances. It is felt that the agreement could be improved somewhat, if the decreasing tendencies of the permittivity and the gyromagnetic ratio (discussed in Section III) could be taken into account accurately. Fig. 5 shows the behavior of μ in the region below magnetic saturation. Since the F_{0n1} resonances shown in Figs. 2 and 3 cover only a limited portion of the X-band range, it was necessary to use the F_{0n1} resonances associated with a cavity of greater height. The measured F_{0n0} and F_{0n1} resonances for a cavity 0.195 inch high are shown in Fig. 6. The apparent absence of F_{0n1} resonances above magnetic saturation of course limits the values of μ and κ to the region below magnetic saturation. The general experimental accuracy may be checked by comparing the F_{0n0} resonant curves in Fig. 6 with those in Fig. 3.

The gyromagnetic ratio γ can be found from

$$|\gamma| = \omega/B^i$$

where ω and B^i are values on the asymptotic line for the F_{0n0} resonant curves. If (14) is differentiated with respect to B^i and the resulting expression for $d\omega_n/dB^i$ expanded in terms of

$$\frac{\Omega_n^2}{2\epsilon_r} \left(\frac{\gamma^2 B^{i^2}}{2} + \frac{\Omega_n^2}{2\epsilon_r} \right),$$

the first term is seen to give

$$|\gamma| \approx \sqrt{\frac{\omega_n}{B^i} \frac{d\omega_n}{dB^i}} \quad (15)$$

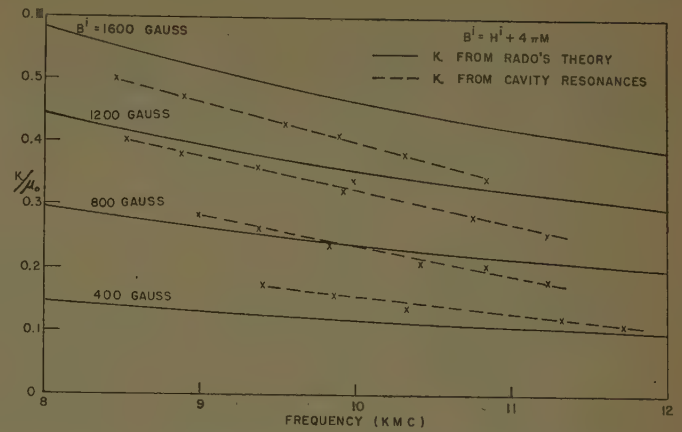


Fig. 4—The off-diagonal tensor permeability component κ for a MgMn-ferrite (General Ceramics R-1 Ferrite).

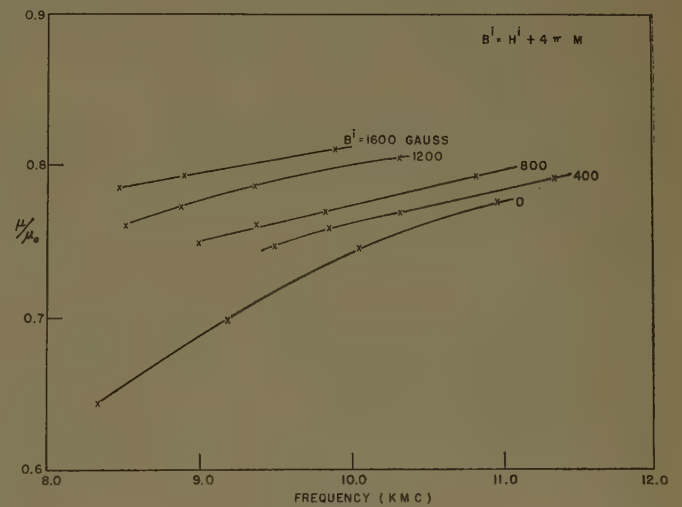


Fig. 5—The diagonal tensor permeability component μ for a MgMn-ferrite (General Ceramics R-1 Ferrite).

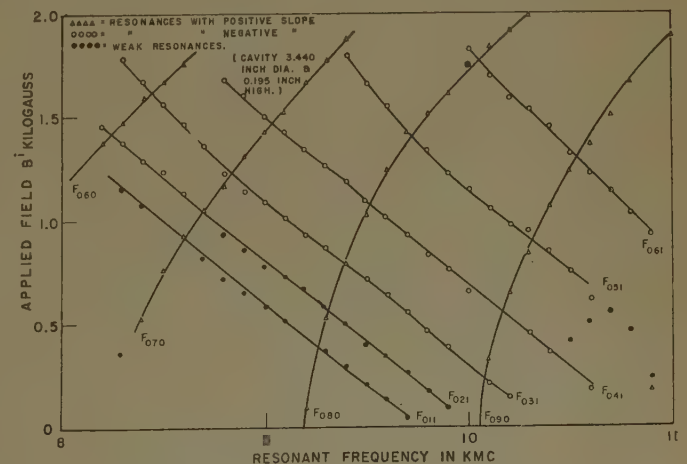


Fig. 6—The measured curves of resonance for another ferrite-filled cavity.

Using the curve for the F_{030} mode in Fig. 3, the value of $|\gamma|$ is found to be 18.5×10^6 rad/oersted-second which compares very well with the value of 18.6×10^6 rad/oersted-second determined from the waveguide measurement described in Section III. The estimated error in using (15) in this instance is 2 per cent.

APPENDIX

THE GRAPHICAL SOLUTION FOR THE F_{0n1} RESONANCES

The calculation of the resonant frequencies from the general characteristic equation for the symmetrical modes was facilitated by a graphical method as illustrated in Figs. 7 and 8. For the F_{0n1} resonances the propagation constant β is π/h , where h is the height of the cavity. At a particular magnetization state of the ferrite, substitution of the relations for the different parameters as discussed in Section II will reduce both sides of (7) to functions of a single independent variable, ω . The F_{0n1} resonances will be determined by the intersections of the curves G_1 and G_2 representing the left and right sides of (7), respectively.

When the ferrite is in the demagnetized state, G_1 and G_2 are identical in magnitude but opposite in sign as illustrated in Fig. 7. The TE and TM modes of resonance in a dielectric filled cylindrical-cavity correspond to the zeros ($J_1=0$) and the poles ($J_0=0$) of the G functions, respectively.

As the ferrite becomes magnetized, the curves G_1 and G_2 will move relative to each other with the curves maintaining essentially the same general shape. The relative motion of the G functions determines the change of the resonant frequencies with respect to the applied static magnetic field. When the magnetization of the ferrite increases beyond certain values, at the intersections such as C_1 , C_2 , C_3 , and C_4 , shown in Fig. 8, x_1 becomes imaginary while x_2 remains real, and the resulting motion of the intersections traces out a family of simple curves which are plotted as the F_{0n1} curves of resonance

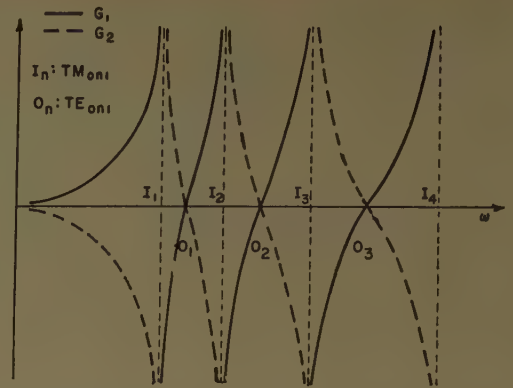


Fig. 7—Sketch of G_1 and G_2 for the demagnetized state.

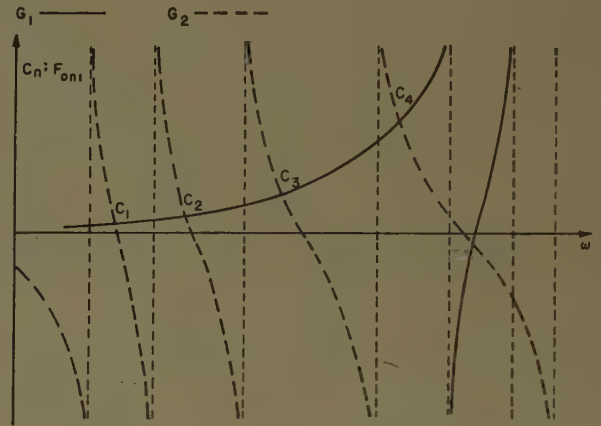


Fig. 8—Sketch of G_1 and G_2 for a magnetized state.

shown in Fig. 2. When the magnetization of the ferrite is below these values, both x_1 and x_2 are real at the intersections, it is not very easy to follow their motion. But it can be shown by careful observation that when the demagnetized state is approached, the F_{0n1} resonances will be reduced to $TM_{0,(n+1)/2,1}$ resonances if n is odd, and will be reduced to $TE_{0,n/2,1}$ resonances if n is even.

Attenuation in Wedge and Septate Waveguides*

ROBIN M. CHISHOLM†

Summary—The wedge waveguide consists of a circular cylinder, a cross section of which has a sector occupied by a metal wedge. As the wedge angle approaches zero the wedge waveguide becomes a septate waveguide. The presence of the wedge permits a mode of lower cutoff frequency than the circular $H_{1,1}$ mode to propagate for a given-diameter waveguide, and the transverse fields associated with this mode become infinite at the tip of the wedge. Because of this, the standard impedance condition, used to calculate the flow of power into the waveguide walls, yields a coupled mode which does not satisfy the Meixner edge condition. This in turn yields extremely high values for the ohmic losses near the tip of the wedge and, as the wedge angle approaches zero, the integrals involved in calculating these losses fail to converge.

In the present paper the results of a careful analysis (presented in an earlier paper) of the behaviour of fields near the tip of a metal wedge are presented. The surface impedance condition is modified to agree with this analysis and used to calculate the attenuation constant of wedge waveguides. Graphs of attenuation constant vs frequency and vs wedge angle are presented for the wedge waveguide operating in its lowest mode. The results of an experimental determination of the attenuation constant are also presented and found to be in agreement with the calculated value. The losses do not become large as the wedge angle is allowed to approach zero and the attenuation constant of the septate waveguide compares favorably with that of standard waveguides.

I. INTRODUCTION

IN this paper an expression is developed for the attenuation constant of the wedge or septate waveguide shown in cross section in Fig. 1. The presence of the wedge or septum (zero angle wedge) means that the standard techniques for calculating the attenuation constant of a waveguide must be used with caution. The difficulties which arise near the tip of the wedge are discussed in Section II and the results of an earlier paper¹ which resolve these difficulties are used to develop a formula for the coupled electric field along the surface of the wedge. In Section III a formula is developed for the attenuation constant of the wedge waveguide for any wedge angle. Since the Bessel functions involved in the formula are not readily available, numerical results are presented graphically giving the attenuation constant as a function of wedge angle, frequency, and wall conductivity. In Section IV an experiment, which was used to check the theoretical attenuation of a particular wedge waveguide is described. Both theory and experiment show the wedge waveguide to be an efficient transmission device.

* This work was supported by a grant extended to the Dept. Elec. Engrg., University of Toronto, Toronto, Can., by the Defence Research Board of Canada, under Extramural Research Grant DRB 5540-02.

† Dept. of Elec. Engrg., Queen's University, Kingston, Ontario, Canada.

¹ R. M. Chisholm, "Ohmic losses in conducting wedges," presented at the Joint URSI-IRE Meeting, Washington, D. C.; May, 1959. To be published shortly.

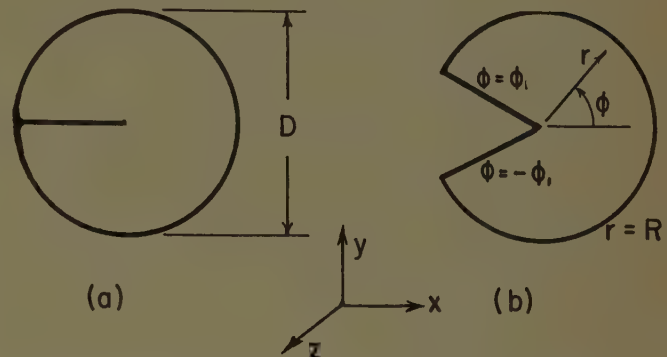


Fig. 1—(a) The septate waveguide; (b) the wedge waveguide.

II. DIFFICULTIES NEAR THE TIP OF THE WEDGE

The standard technique used in calculating the attenuation constant of a waveguide assumes that the tangential magnetic field at the walls of the waveguide is the same as it would be if the walls were perfectly conducting. A coupled electric field, tangent to the surface and at right angles to the tangential magnetic field, is then assumed to be proportional to this tangential magnetic field. In the wedge waveguide this would mean an axial electric field E_z^0 on the surface of the wedge proportional to H_r^0 , the radial magnetic field associated with perfectly conducting walls. The lowest propagating mode in such a waveguide, however, is governed by the axial magnetic field,

$$H_z^0 = J_t(\alpha r) \sin(t\phi) \quad (1)$$

where t lies between $\frac{1}{2}$ (for the zero angle wedge) and unity (for the 180° wedge). The radial magnetic field, in this case, is given by

$$H_r^0 = (-j\beta/\alpha) J_t'(\alpha r) \sin(t\phi) \quad (2)$$

and becomes infinite as r^{t-1} when r approaches 0. The axial electric field coupled to H_r^0 by the finite wall conductivity would, therefore, become infinite at the tip of the wedge in violation of the Meixner edge condition.

The power loss within the wedge, moreover, is found by integrating the normal component of the Poynting vector formed by the tangential magnetic field and the coupled electric field on the surface of the wedge. If both H_r and E_z become infinite at the tip of the wedge very large contributions come from the tip region. In particular, if the wedge angle is allowed to approach zero, the Poynting vector behaves as r^{-1} near the tip of the wedge and its integral over the face of the wedge does not converge at all.

In a recent paper¹ by the author this problem was discussed by extending the power series approach used by

Meixner² for dielectric wedges, and by examining carefully the surface impedance concept as applied to wedges. The results of that paper may be summed up as follows.

1) Near the tip of the wedge static boundary conditions must be satisfied and, as r approaches zero,

- a) E_r and E_ϕ behave as $r^{t(E)-1}$ or as $r^{t(H)+1}$, while
- b) H_r and H_ϕ behave as $r^{t(H)-1}$ or as $r^{t(E)+1}$.

The discrete values which $t(E)$ and $t(H)$ can assume depend on the properties of the wedge material and on the wedge angle. The exponent $t(E)$ depends only on the electric properties of the wedge while $t(H)$ depends only on the magnetic properties of the wedge. These eigenvalues can be arranged in order of increasing magnitude in problems involving ordinary metal wedges. For *non-ferrous* metals the smallest allowable value of $t(H)$ lies very close to unity while the smallest allowable value of $t(E)$ is less than $t(H)$ and lies very close to the value of t used in the "perfectly conducting" expression for the lowest mode in (1). Since the lowest mode appears to be governed by the lowest eigenvalue of $t(E)$, the transverse fields H_r and H_ϕ , associated with this lowest mode, are also governed by $t(E)$ and must, therefore, approach zero as $r^{t(E)+1}$ at the tip of the wedge.

2) The surface impedance condition, moreover, holds on the faces of a metal wedge to within a few skin depths of the tip of the wedge. It does not appear to break down in a violent manner either, when it is assumed to hold over the entire wedge face.

3) The apparent breakdown of the surface impedance concept in the wedge waveguide problem arises from the assumption that, on the faces of the wedge,

$$E_s^e = \pm Z_s H_r^0 \quad (3)$$

where H_r^0 is the radial magnetic field associated with the perfectly conducting wedge, and Z_s is the surface impedance of the wedge material which is related to the conductivity, σ , of the metal by the relationship

$$Z_s = (j\omega\mu/\sigma)^{1/2}.$$

It is a simple matter to show that the coupled electric field E_s^e has associated with it a radial magnetic field given by

$$H_r^e = (j\omega\epsilon/\alpha^2)(1/r) \frac{\partial E_s^e}{\partial \phi} \quad (4)$$

in which $\alpha^2 = k^2 - \beta^2$ where $\beta = (2\pi/\lambda_0)$ and $k = (2\pi/\lambda_0)$. As r approaches zero, therefore, the ratio of H_r^e to H_r^0 behaves as

$$\frac{H_r^e}{H_r^0} = \frac{(KZ_s)}{r} \quad (5)$$

For any finite r , this ratio approaches zero as the surface impedance, Z_s , approaches zero (conductivity becomes infinite), but, for fixed Z_s , the ratio becomes infinite as r approaches zero. This means, therefore, that the surface impedance condition should be written in the form

$$E_s^e = Z_s(H_r^0 + H_r^e). \quad (6)$$

This is a differential equation for E_s^e if H_r^0 is a known function, as it is in the wedge waveguide problem, and this differential equation must be satisfied on the faces of the wedge. This equation, moreover, agrees with the power series approach used to examine the behavior near the tip of the wedge. The coupled E_s^e which satisfies (6) can be found approximately and, along the faces of the wedge in the wedge waveguide; it has the form

$$E_s^e = \frac{-\left(\frac{\beta\alpha}{\omega\epsilon}\right)rJ_t'(\alpha r)\left[\left(\frac{r}{bZ_s}\right) + t \cos(t\phi_1)\right]}{\left(\frac{r}{bZ_s}\right)^2 + t^2} \quad (7)$$

where $b = (j\omega\epsilon/\alpha^2)$ and $\phi = \pm\phi_1$ are the faces of the wedge in the coordinate system shown in Fig. 1. Assuming the total, radial magnetic field, H_r , to be proportional to the E_s^e given by (7), a typical waveguide has a radial magnetic field which, near the tip of the wedge, behaves in the manner shown in Fig. 2.

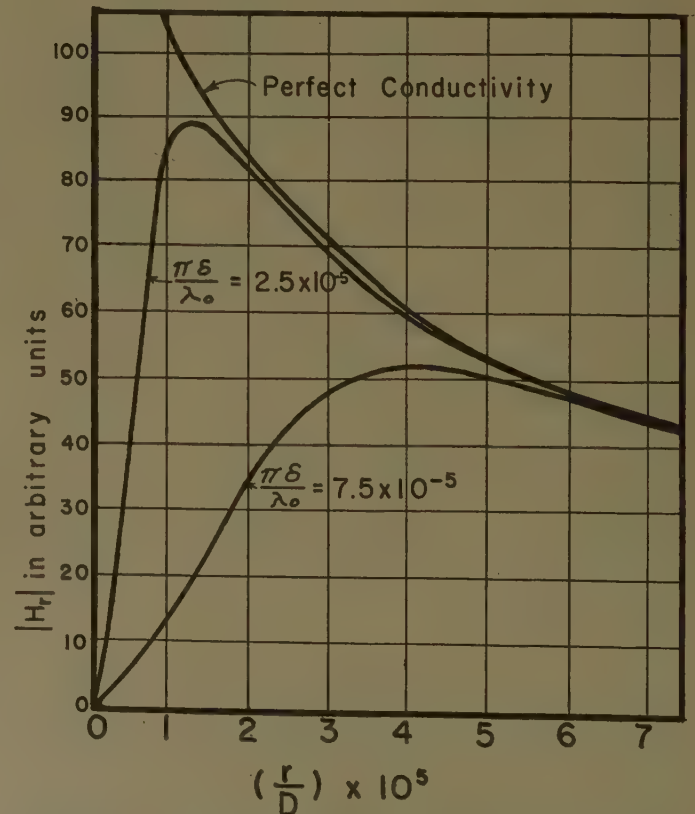


Fig. 2—The variation of the total radial magnetic field with distance from the tip of the wedge in a 1° wedge waveguide with a free-space wavelength to diameter, D , ratio of 2.0.

² J. Meixner, "The Behaviour of Electromagnetic Fields at Edges," New York University, New York, N. Y., Rept. No. EM-72.

As the conductivity becomes infinite the peak of the curve moves to the left and upward, approaching the "perfectly conducting" curve in the same manner as that in which a finite Fourier series approaches a discontinuous function as more and more terms are taken. It always reaches a maximum, however, and vanishes at $r=0$ for all values of conductivity.

III. THE ATTENUATION CONSTANT

Using the modified value for E_z^c given by (7), the loss calculations are quite straight forward. In terms of rms field values the power flow into the wedge is given by

$$P_w = 2 \operatorname{Re} (\tilde{Z}_s)^{-1} \int_0^R (E_s^c \tilde{E}_s^c + E_r^c \tilde{E}_r^c) dr \quad \text{watts per meter.} \quad (8)$$

In this equation, $E_r^c = \pm Z_s H_z^0$ for all values of r since H_z^0 vanishes for small values of r and no violation of the edge condition results. As in a circular waveguide, the power lost to the circular periphery, $r=R$, is given by

$$P_p = \operatorname{Re} (\tilde{Z}_s)^{-1} \cdot \int_{-\phi_1}^{\phi_1} (|E_s^c|^2 + |E_\phi^c|^2) R d\phi \quad \text{watts per meter.} \quad (9)$$

The attenuation constant, γ , is given by

$$\gamma = \frac{P_w + P_p}{2P_s} \quad \text{nepers per meter} \quad (10)$$

where P_s is the power flow along the axis of the waveguide given by

$$P_s = \operatorname{Re} \int_{-\phi_1}^{\phi_1} \int_0^R (E_r \tilde{H}_\phi - E_\phi \tilde{H}_r) r dr d\phi \quad \text{watts.} \quad (11)$$

Since the Bessel functions involved in these calculations are not readily available in tabulated form it is convenient to express the final formula for the attenuation constant in a form which is easy to handle on an electronic computer, which has to handle the functions involved using the power series method. With this end in view the attenuation constant given by (10) can be written as the sum of two terms, γ_w arising from the wedge losses, and γ_p arising from the periphery losses. Writing

$$K_1 = (z_1)^4 \pi^{-2} (\pi \delta / \lambda_0) (\lambda_0 / D) (\lambda_g / D) \quad (12)$$

where

- z_1 = first root of $J'_1(z) = 0$,
- δ = skin depth of the metal,
- λ_0 = free-space wavelength,
- λ_g = guide wavelength, and
- D = the diameter of the waveguide;

γ_w can be written in the form,

$$\gamma_w D = \frac{(4K_1 I_1)}{\phi_1 I_2} \quad \text{nepers per diameter,} \quad (13)$$

and γ_p can be written in the form

$$\gamma_p D = \frac{K_1 F_2(1/2) [(\pi t D / Z_1^2 \lambda_g)^2 + 1]}{I_2} \quad \text{nepers per diameter.} \quad (14)$$

The integrals I_1 and I_2 and the function $F_2(x)$ are defined in the Appendix. These functions lend themselves quite readily to numerical evaluation using an electronic computer to evaluate the fractional order Bessel functions and to perform the numerical integration. The Ferranti Mark I electronic digital computer, formerly at the University of Toronto and presently at the National Research Laboratories in Ottawa, Ontario, has been coded to evaluate (13) and (14) as functions of wedge angle and the normalized parameters, (λ_0/D) and $(\pi \delta / \lambda_0)$. The results are shown in Figs. 3, 4, and 6.

Fig. 3 shows the way in which the attenuation constant, γ , varies with wedge angle for three different skin depth ratios. For brass at 1000 mc ($\pi \delta / \lambda_0$) is approximately 4×10^{-5} . As the wedge angle approaches zero γD approaches a finite value.

Fig. 4 shows the way in which the attenuation constant varies with frequency for a constant skin-depth ratio of 5×10^{-5} using a 1° wedge. Since γD varies almost directly with the skin-depth ratio, $(\pi \delta / \lambda_0)$, and since the variation in γD with wedge angle is small, Fig. 4 can be used to find an approximate value for γ for any wedge angle less than 20° and any reasonable conductivity, without introducing an error of more than 10 per cent in the final answer. The high-frequency conductivity of metals is usually the unknown factor in practice which prevents accurate attenuation calculations and more detailed loss curves are usually not required.

IV. EXPERIMENTAL WORK

The calculations were checked experimentally using a shorted, half-wave length of wedge waveguide as a resonant cavity. The dimensions of the cavity used are shown in Fig. 5. The Q of the cavity was measured using it as a two-line cavity coupling system.³ Two variable length probes were used and a plot of Q vs probe length was extrapolated to obtain an estimation of the unloaded Q which the cavity would have if both probes were removed. The cavity was resonated in its lowest mode at 602.1 mc and a half-power bandwidth varying between 0.06 and 0.09 mc was obtained.

Fig. 6 shows the theoretical variation (solid curve) of γD with the skin-depth ratio, $(\pi \delta / \lambda_0)$, for an 18° wedge with the same wavelength to diameter ratio used in the experiment. The skin depth ratio for brass at 600 mc,

³ C. G. Montgomery, "Techniques of Microwave Measurements," MIT Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., no. 11, 1st ed., p. 289; 1947.

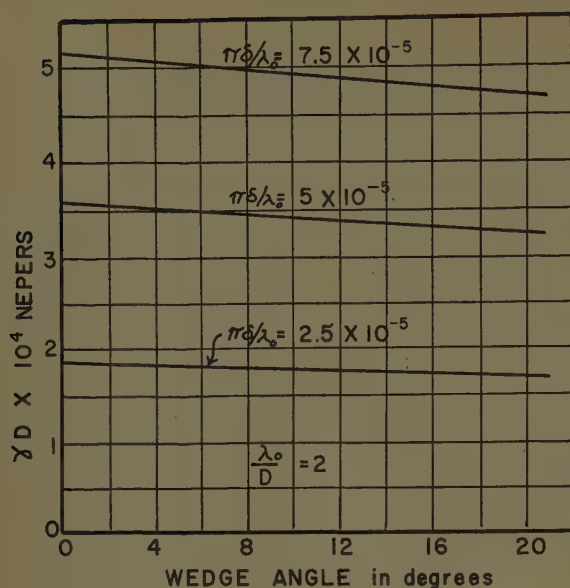


Fig. 3—The variation of the attenuation constant γ with wedge angle in nepers per diameter.

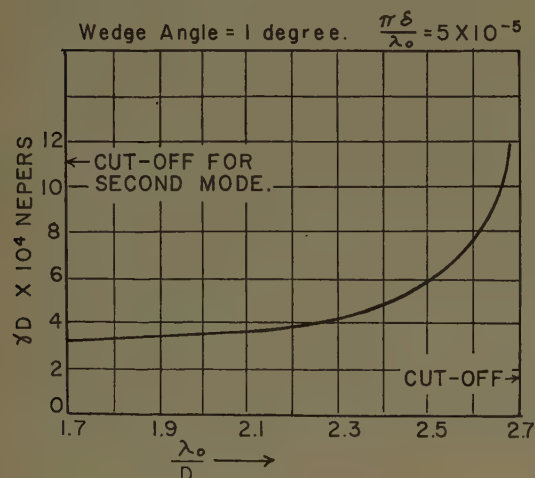


Fig. 4—The variation of the attenuation constant γ with free-space wavelength in nepers per diameter.

based on dc values of conductivity, lies somewhere between 2.5×10^{-5} and 3.2×10^{-5} . Using the curve in Fig. 6 together with the formula

$$Q = \frac{\pi(\lambda_g/\lambda_0)^2}{(\gamma D)(\lambda_g/D) + 4(\lambda_0/\lambda_g)(\pi\delta/\lambda_0)} \quad (15)$$

which relates the cavity Q to the waveguide attenuation constant of the waveguide forming the cavity; a skin-depth ratio between 3.3×10^{-5} and 4.9×10^{-5} would be required to yield the experimental results obtained. This is in keeping with the fact that the RF conductivity of metals is considerably lower than the dc conductivity as a result of unfavourable surface conditions.

As a further check on the over-all calculations the same computer program was used to calculate the attenuation constant of the 18° wedge waveguide operating in its second-lowest mode with a wavelength to diameter ratio equal to that of the second-lowest reso-

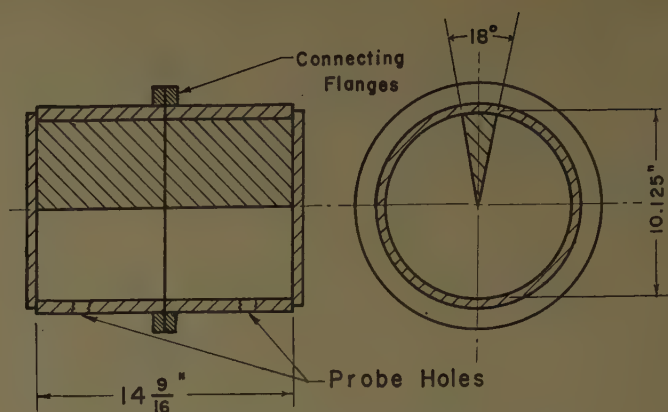


Fig. 5—A shorted half-wavelength wedge waveguide used as a resonant cavity.

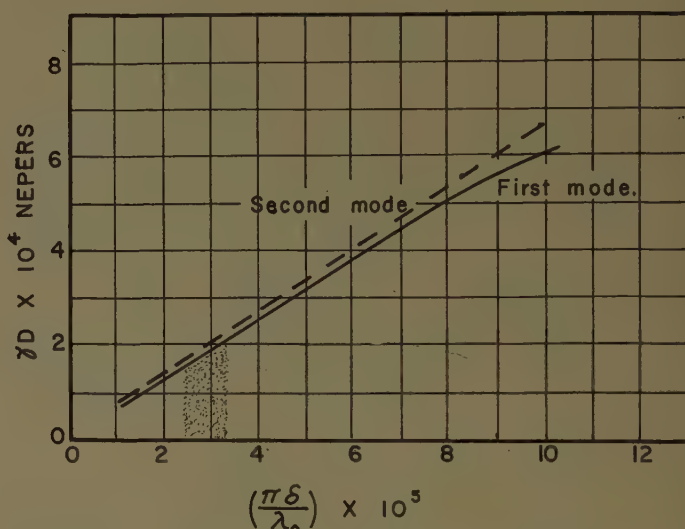


Fig. 6—The variation of the attenuation constant γ with the skin-depth ratio for the 18° wedge waveguide at the wavelength to diameter ratios used in the experiment. The ratio of free-space wavelength to diameter is 1.94 for the lowest mode and 1.43 for the second-lowest mode. The shaded region indicates the expected skin-depth ratio for brass at 600 mc based on the dc conductivity.

nant frequency of the cavity in Fig. 5. The cavity was then resonated in its second-lowest mode at a frequency of 818.0 mc and a half-power band width between 0.075 and 0.135 mc was obtained. Assuming the skin depth ratio, $\pi\delta/\lambda_0$, to vary as the square root of the frequency, this would require from the dotted curve in Fig. 6, at 600 mc and the same ratio of free-space wavelength to diameter as that used in the experiment, a skin-depth ratio between 4.8×10^{-5} and 8.7×10^{-5} to yield the experimental results obtained. The second mode is almost the same as the dominant $H_{1,1}$ circular waveguide mode. None of the field components become infinite at the tip of the wedge so that the special wedge theory discussed earlier does not apply and standard methods of calculating the attenuation constant are known to be valid. Because of this it was originally intended that this mode could be used to establish the correct skin-depth ratio for the cavity under test. Since the experiment was designed, however, it has been noted that the second mode lies quite close to the waveguide cutoff at the frequency

used in the experiment. In this frequency range the attenuation constant varies rapidly with frequency and, as a result of this, difficulties arose in measuring the Q of the second mode which did not arise in measuring the lowest mode Q . This accounts for the large uncertainty in the second-mode results and suggests that more weight should be put on the agreement between the results of the lowest-mode experiment and the skin-depth ratio based on dc values.

Since the second-lowest mode is very close to the $H_{1,1}$ circular waveguide mode, the dotted curve in Fig. 6 should yield an attenuation constant for brass pipe which lies very close to the value found from standard circular waveguide curves. Taking into account the fact that the wedge waveguide is closer to its cutoff frequency than the circular waveguide for the same free-space wavelength to diameter ratio, there is very good agreement between these results and those found in textbooks.⁴

It should also be pointed out that about 63 per cent of the losses in either a wedge or septate waveguide come from the wedge or septum. By making the wedge or septum, which require no physical strength, out of copper instead of brass a further improvement can be obtained.

V. CONCLUSIONS

In conclusion one can say that wedge and septate waveguides are efficient transmission devices with attenuation constants comparable to those of conventional waveguides. They have a lower cutoff frequency than the circular waveguide for a given diameter and the dominant modes are stable. The high power losses in the wedge, which occur as a result of the high field strengths near the tip of the wedge, are compensated for by the high power transfer associated with these high field strengths.

The fact that the radial magnetic field must vanish at the tip of the wedge, as shown in Fig. 2, indicates

that the exact nature of the tip region is not critical. Making the tip very sharp does not increase the losses beyond a certain limit and rounding the tip off does not necessarily decrease the losses since high surface currents would then flow over a larger area.

Curves have been presented which enable the calculation of the attenuation constants of wedge and septate waveguides under varying conditions and these curves agree reasonably well with experimental evidence. The mechanism is available, moreover, for calculating the attenuation constants of such waveguides to a higher degree of accuracy than that obtainable by using the curves presented in Section III, if such accuracy is required.

APPENDIX

The quantities used in (13) and (14) are defined as follows.

$$I_1 = \int_0^{1/2} [F_1(x) + F_2(x)] dx. \quad (16)$$

$$I_2 = \int_0^{1/2} [G_1(x) + G_2(x)] dx. \quad (17)$$

$$F_1(x) = 2 \left(\frac{\lambda_0}{\lambda_g} \right)^2 \left(\frac{\lambda_0}{\pi \delta} \right)^2 (x \Gamma(t) J_t'(2z_1 x))^2 (z_1)^{2-2t} \times \frac{(k_2 x^2 - k_3 x + k_4)}{(k_2 x^2)^2 + t^4}. \quad (18)$$

$$F_2(x) = (\Gamma(t) J_t(2z_1 x))^2 z_1^{-2t}. \quad (19)$$

$$G_1(x) = x^{-1} (t \Gamma(t) J_t(2z_1 x))^2 (Z_1)^{-2t}. \quad (20)$$

$$G_2(x) = 4 Z_1^{2-2t} x [\Gamma(t) J_t'(2z_1 x)]^2. \quad (21)$$

$$K_2 = 2 \left(\frac{\lambda_0}{\pi \delta} \right)^2 \left[1 - \left(\frac{\lambda_0}{\lambda_g} \right)^2 \right] (Z_1)^2. \quad (22)$$

$$K_3 = 4 \left(\frac{\pi \delta}{\lambda_0} \right) t z_1 \left[1 - \left(\frac{\lambda_0}{\lambda_g} \right)^2 \right]^{1/2} \tan(t\pi). \quad (23)$$

$$K_4 = 4 \left(\frac{\pi \delta}{\lambda_0} \right)^4 \tan^2(t\pi) t^2. \quad (24)$$

⁴ E. C. Jordan, "Electromagnetic Waves and Radiating Systems," Prentice-Hall, Inc., New York, N. Y., 1st ed., p. 290; 1950.

Diffraction of Nearly Plane 3.2-cm EM Waves by 45° and 90° Conducting Wedges. Comparison with Theory*

N. E. HEDGECOCK† AND A. B. McLAY‡

Summary—Diffraction of 3.2-cm microwaves has been investigated when 45° and 90° conducting wedges were placed, in turn, in a beam with the diffracting edge on the axis of propagation and 5.5 meters from the apex of a pyramidal horn radiator, and with the beam incident on the forward "bright" surface of each wedge normally, and also at 45°. The incident field was polarized with the electric vector parallel to a diffracting edge. Field patterns for each case out to 7.5λ (24 cm) from an edge have been observed along a line at right angles to the axis of propagation and on seven parallel lines at wavelength intervals further from the source. Comparison of measured intensities with those calculated with a Bendix G15D computer has been made using asymptotic formulas derived by Pauli.¹ Comparison with values obtained from the exact formula of Macdonald² has also been made in a limited region close to a diffracting edge only, since the formula converges very slowly except within some two to three wavelengths distance from the edge. Agreement between experiment and theory is quite good in general in the central region of the field studied. Discrepancies that occur in certain regions have been accounted for.

INTRODUCTION

THE experimental results reported here on diffraction by conducting wedges were obtained in the course of an extensive and still-continuing investigation on the microwave optics of a large 45°-90°-45° dielectric prism, with and without an aluminum foil coating. The prism dimensions are 34 cm × 34 cm × 48 cm × 50 cm height. When coated, it conveniently provided two 45° wedges and a 90° wedge, each with finite bounding surfaces, but large enough to be considered as practically semi-infinite planes. The conducting covering was household aluminum "foil-wrap." The skin penetration depth of aluminum is of the order of 10^{-5} wavelengths at $\lambda = 3.2$ cm, so the foil-covered prism is essentially a perfect conductor. The general arrangement of apparatus for the free-field diffraction measurements and details of circuitry, field probe motor drive, and automatic synchronization with recorder chart drive have been described previously by Jordan and McLay.³ The prism in each orientation used was placed with the center of the particular diffracting corner edge on the axis of propagation of the pyramidal horn radiator and at a distance of 5.5 meters from the projected apex of the horn in the H plane. The field probe was a 1N415-B crystal which approximates in length a simple

Hertzian dipole antenna and has nearly square-law response for the signals observed. Probe runs were made from each diffracting edge out on a line at right angles to the axis of propagation and on parallel lines 1.6 cm apart, farther from the source, leading out from the inside the geometrical shadow edge. In this paper, results are given for observations made at 3.2 cm intervals only, since those of alternate runs do not contribute any additional significant detail and tend to confuse the plotted field patterns.

In all experiments the incident radiation was polarized with the electric vector parallel to the edge of the wedge, to which observations were made in a plane of incidence perpendicular to it. The diffraction problem is therefore a two-dimensional scalar one.

A number of theoretical treatments of conducting wedge diffraction, including some recent ones, have been reported since Sommerfeld (1894-1896) first proposed a general method of solution. A very good summary, with bibliography, is included in a paper by Oberhettinger⁴ on conducting wedges in plane, cylindrical and spherical incident waves. This paper appeared as we were completing computations, using plane incident wave solutions of Macdonald² and Pauli.¹ The asymptotic solutions of Pauli were used, rather than the somewhat more exact ones of Oberhettinger⁵ and Felsen,⁶ because of greater ease of computation.

There has been very little experimental investigation of diffraction by conducting wedges, except for the special case of a semi-infinite thin sheet (wedge of angle zero). In the microwave region Watson and Horton⁷ observed the field on a circle of radius 12λ (≈ 36 cm) about the edge of a 22.5° wedge. The radiation was incident in a direction perpendicular to the plane of the back (shadow) surface. Results showed fair agreement with calculated ones using Pauli's¹ asymptotic solution. Row⁸ observed the field from the geometrical shadow boundary out at a distance of 4λ behind the edge of a 0°, a 45° and a 90° wedge in a 3.185-cm cylindrical radiation incident normally on the front (illuminated) surface. Re-

* This investigation has been carried out with aid of a National Research Council of Canada research grant.

† Max Planck Institut für Chemie, Mainz, West Germany. Formerly with McMaster University, Hamilton, Ontario, Can.

‡ McMaster University, Hamilton, Ontario, Can.

¹ W. Pauli, "On asymptotic series for functions in the theory of the diffraction of light," *Phys. Rev.*, vol. 54, pp. 924-931; December 1, 1938.

² H. M. Macdonald, "Electric Waves," Cambridge University Press, London, Eng., Appendix D; 1902.

³ C. E. Jordan and A. B. McLay, "Diffraction of 3.2 cm electromagnetic waves by dielectric rods," *Can. J. Phys.*, vol. 35, pp. 1253-1264; November, 1957.

⁴ F. Oberhettinger, "On the diffraction of waves and pulses by wedges and corners," *J. Res. Natl. Bur. Stand.*, vol. 61, pp. 343-365; November, 1958.

⁵ F. Oberhettinger, "On asymptotic series for functions occurring in the theory of diffraction of waves by wedges," *J. Math. Phys.*, vol. 34, pp. 245-255; January, 1956.

⁶ L. B. Felsen, "Alternative field representations in regions bounded by spheres, cones and planes," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP5, pp. 109-129; January, 1957.

⁷ R. B. Watson and C. W. Horton, "On the diffraction of a radar wave by a conducting wedge," *J. Appl. Phys.*, vol. 21, pp. 802-804; August, 1950.

⁸ R. V. Row, "Microwave diffraction measurements in a parallel plate region," *J. Appl. Phys.*, vol. 24, pp. 1448-1452; December, 1953.

sults showed fair agreement with those calculated using an equation similar to Pauli's but with an approximate correction for a line source at only 35.7λ from the edge. In our investigation the fields have been observed in some detail from 0 to 7.5λ behind each wedge corner and from the unilluminated surface out to 8λ from the geometrical shadow boundary for 45° and 90° wedges with normal and 45° incidence. Comparison with theory has been made over the whole region investigated.

RESULTS

The results of field measurements and calculations are given in the plots in Figs. 1-4 of relative optical intensity (observed $I/\text{incident } I_i$) against position y taking the axis of propagation as x axis of the experimental field position geometry. The average value of I_i is shown by a vertical T line opposite the curves for $x=0$. Each experimental curve represents the average obtained from two sets of measured values. One set was

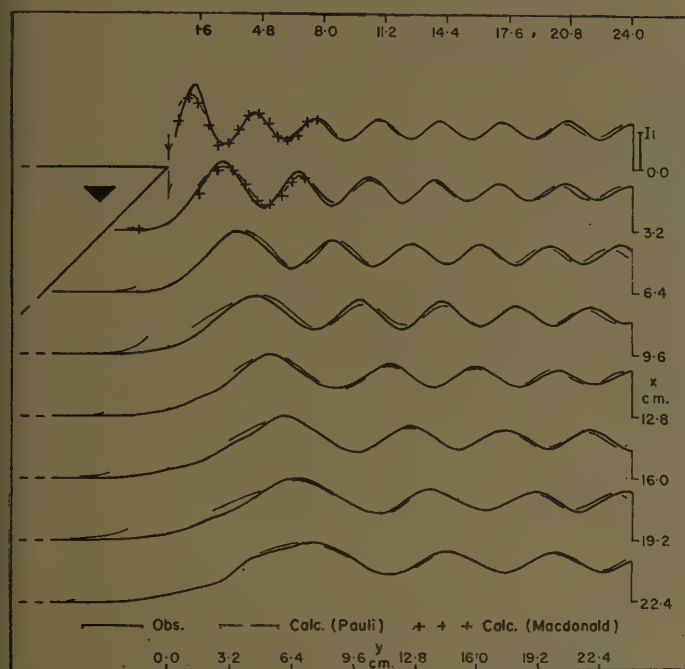


Fig. 1—Diffraction pattern of a 45° conducting wedge in 3.2-cm radiation, incident normally on a 48-cm wide face. Intensity $I = |E_x|^2$ vs position y .

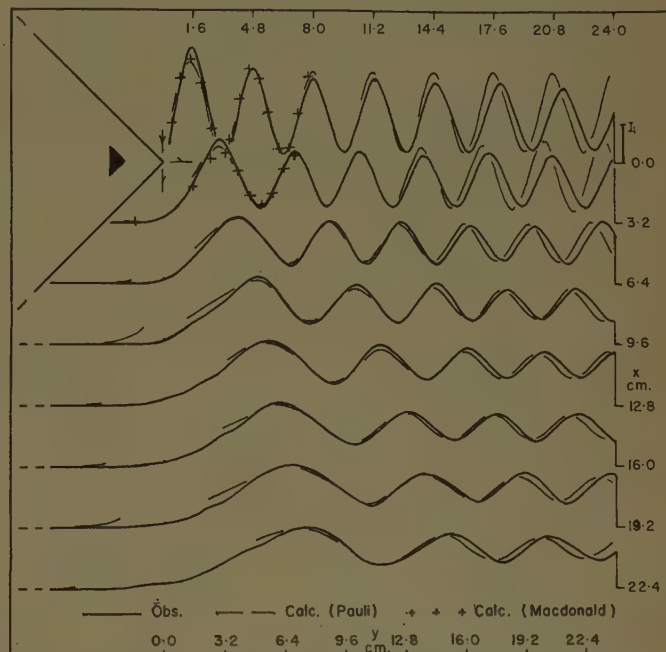


Fig. 3—Diffraction pattern of a 90° conducting wedge in 3.2-cm radiation incident at 45° . Intensity $I = |E_x|^2$ vs position y .

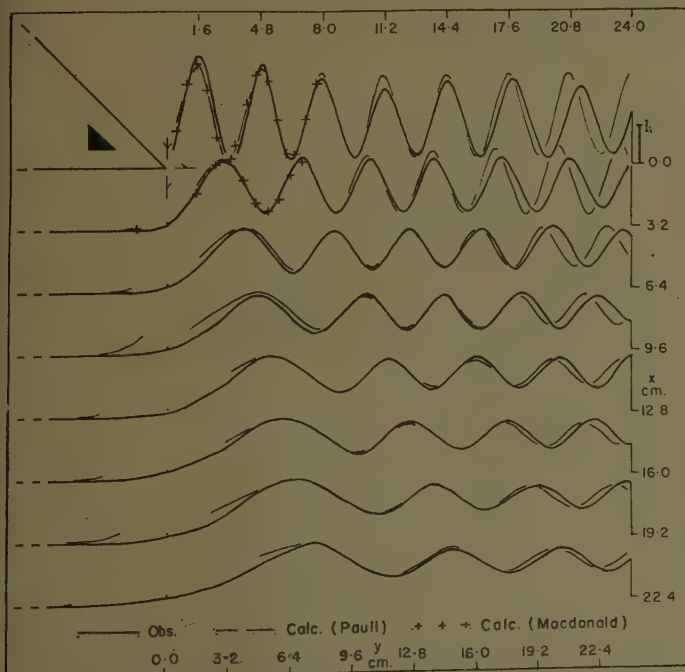


Fig. 2—Diffraction pattern of a 45° conducting wedge in 3.2-cm radiation, incident at 45° on a 48-cm wide face. Intensity $I = |E_x|^2$ vs position y .

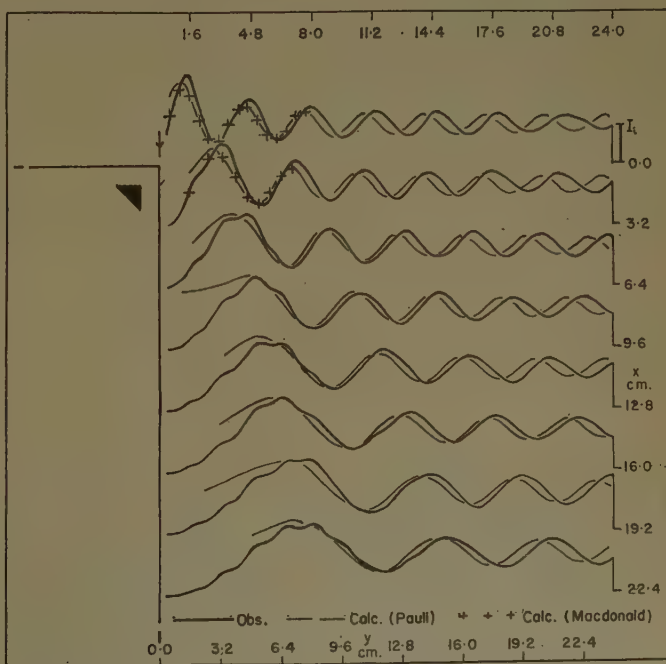


Fig. 4—Diffraction pattern of a 90° conducting wedge in 3.2-cm radiation, incident normally. Intensity $I = |E_x|^2$ vs position y .

made with geometry corresponding exactly with that given in the appropriate figure, in which the prism orientation is shown by the small inset solid triangle. The other set was made with mirror image geometry about the x axis. Experimental patterns were also obtained for a 45° wedge with incidence normally and at 45° on a 34-cm wide prism face. These were so similar to the patterns of Figs. 1 and 2, respectively, that they are not included here.

Computation of fields was greatly facilitated by the use of a Bendix G-15D General Purpose Digital Computer. The $+$ symbols in the field plots give relative intensities I/I_i , calculated by using the Bendix Intercom 103 programming, and an exact formula of Macdonald (1902) for the amplitude (normalized for an incident plane wave of unit amplitude, and with time dependence $\exp(i\omega t)$ omitted).

$$E_s = \left(\frac{4\pi}{\alpha}\right) \sum_{n=1}^{\infty} J_p(kr) \sin p\theta \sin p\theta_1 e^{ip\pi/2}. \quad (1)$$

$|E_s|^2$ is identified with the intensities of the field plots. Here the edge of the wedge is taken along the z axis and the two bounding surfaces are defined by the planes $\theta=0$ and $\theta=\alpha$ (exterior angle of wedge) in cylindrical polar coordinates. $\theta=\theta_1$ is the azimuth of the radius vector from the origin at the wedge edge to the source, $k=2\pi/\lambda$ and $p=n\pi/\alpha$. For the wedges under investigation, in particular internal angles 45° and 90° , $\alpha=7\pi/4$ and $3\pi/2$, respectively. In the two cases of incidence used for each wedge, $\theta_1=\pi/2$ (normal) and $\theta_1=\pi/4$. The series of (1) converges fairly rapidly for small kr , but its evaluation becomes laborious and subject to accumulation of errors for kr larger than the maximum value 15 used in computation.

The broken thin lines in Figs. 1 to 4 are curves obtained from computations using Bendix Intercom 1000 programming and asymptotic equations of Pauli.¹ For the particular region of field and state of polarization investigated here,

$$E_s = v(r, \theta - \theta_1) - v(r, \theta + \theta_1), \quad (2)$$

where

$$v(r, \phi) \simeq e^{ikr \cos \phi} + (2\pi kr)^{-1/2} e^{-i(kr + \pi/4)} \frac{\sin \frac{\pi^2}{\alpha}}{\frac{\pi}{\alpha} \frac{\pi^2}{\cos \frac{\pi^2}{\alpha}} - \cos \frac{\pi}{\alpha} \phi}.$$

Here the first term gives the contribution of the incident and reflected fields based on simple geometrical optics with $\phi=\theta-\theta_1$ for the incident and $\phi=\theta+\theta_1$ for the reflected part. The second term gives the contribution associated with the purely diffracted field resulting from incidence and reflection according to the applicable values of ϕ above. The term $e^{ikr \cos \phi}$ has its finite value for $-\pi < \phi < \pi$, but is to be considered equal to zero

otherwise. The equation is not expected to be good for small values of kr , and certainly not good for $\phi \rightarrow \pm\pi$, because of the denominator in the second term of the expression for $v(r, \phi)$.

Pauli derived a special asymptotic solution for evaluation exactly on geometrical optics boundaries, $\phi=\pi$. According to this,

$$v(r, \pi) = \left[\frac{1}{2} + \frac{i}{2} (2kr)^{-1/2} \frac{\pi}{\alpha} \cot \frac{\pi^2}{\alpha} \right] e^{-ikr}. \quad (3)$$

This equation was used for evaluation along the geometrical shadow boundary where $\phi=\theta-\theta_1=\pi$. The calculated curves on the reflected beam boundary, $x=0$, where $\phi=\theta+\theta_1=\pi$ in each case when $\theta_1=\pi/4$, were also obtained using (3). For the normal incidence cases, $\theta_1=\pi/2$, the beam is reflected back towards the source into a region not investigated.

DISCUSSION OF RESULTS

Comparison of the experimental and theoretical results shown in Figs. 1 to 3 indicates good agreement in the central region of each field. There are, however, noticeable discrepancies in certain regions, namely 1) near geometrical optics beam boundaries, 2) at large y values and 3) near a wedge surface. These cases are discussed in the following.

1)—The Pauli¹ field values deviate from those of Macdonald² as kr approaches 0 because of the asymptotic form of his equations. Those calculated using (2) are also unreliable as ϕ approaches $\pm\pi$ because of the form of its denominator. This is evident in the upward deviation from the experimental curves of the Pauli curves both from outside and inside the geometrical shadow edge, $y=0$, $\phi=\theta-\theta_1=\pi$ and in the increase of the Pauli peaks in Figs. 2 and 3 along $x=3.2$ cm where angular closeness to the reflected beam boundary, $x=0$, $\phi=\theta+\theta_1=\pi$, is being approached with increasing y . The Pauli values along $x=0$ in Figs. 2 and 3 and along $y=0$ in Figs. 1 to 3 were obtained using (3) and should be good except for small kr .

2)—The experimental peaks deviate outward from the theoretical ones at larger values of y in all cases, by a greater amount for 45° incidence than for normal. The deviations are closely of order expected, taking into consideration that the experimental incident waves are not plane as assumed in the theory, but arcs of 550–572.4 cm over the range $x=0$ –22.4 cm.

3)—The large peak nearest to the wedge corner, at about $y=\lambda/2$, in each experimental curve at $x=0$ is higher and farther from the corner than the exact theoretical one. The deviation is more noticeable in Figs. 1 and 4, where the theoretical position is closer to the corner, than in Figs. 2 and 3. Also, the first and second peaks away from the wedge in the curve at $x=3.2$ cm in Figs. 1 and 4 deviate noticeably from the theoretical ones. That these discrepancies result from scattering of radiation from the diode probe to the adjacent wedge

surface with rescattering to the probe is confirmed by the appearance of a noticeable standing-wave pattern in the curves of Fig. 4, with partially or unresolved maxima at half-integral values of the wavelength from the surface in the y direction. The effect is most pronounced at medium values of x where backscattering from the surface should be maximum. Some distortion of the first peak at large values of x , in Figs. 1 and 3 may also result from this effect.

The conclusion, from the qualitative treatment of discrepancies between the observed and theoretical results, is that there would be good agreement, except when close to a strongly reflecting surface, if quantitative account were taken of the difference between the finite distance of the experimental source and the infinite distance of the theoretical one, and if exact equations, or more exact asymptotic ones had been used for computation. The semi-infinite wedge surfaces of theory seem to be quite well approximated by the large finite actual surfaces, since there is no significant difference between the experimental results of Fig. 2, for which the strongly reflecting surface has a width of 48 cm, and results obtained when its width was 34 cm (observed

but not reported here). A small but observable modulation of peaks, particularly in the curve for $x=0$, however, may be the result of diffraction by the finite reflecting aperture. This is also noticeable in Fig. 3 for the 90° wedge, with the whole pattern almost identical to that of Fig. 2 above. The calculated values for these 45° and 90° wedge cases are so nearly the same, except for very small kr values, that differences are too small to show at the scale used in the figures. Thus, the fact that the "shadow" surfaces are inclined differently has very little effect.

The experimental defect of probe disturbance is more serious than previously suspected. That much of it is a resonant one can be deduced from the standing wave effects seen in Fig. 4. Resonant effects cannot be eliminated entirely, but experiments are now underway to find a probe with considerably reduced disturbing effect for use in future studies of diffraction of microwaves by conducting objects. The disturbance by the probe used here is probably negligible when using dielectric objects and no such effects have been detected in studies of the microwave optics of the large dielectric prism or of other dielectric objects.

The Matching of Parallel Dielectric Plates to Free Space

G. C. McCORMICK†

Summary—An array of parallel dielectric plates of uniform thickness and separation constitutes an anisotropic medium which is useful as a polarizer. Reflections can be minimized by the use of quarter-wave steps at the two faces. A weather protecting skin is also of practical interest. The phase shifts at the boundary discontinuities form the subject of this paper. While in some instances these phase shifts are of marginal significance, nevertheless they cannot be ignored in the design of a polarizer of the highest performance, or if the use of materials having a high dielectric constant is considered. The scattering coefficients are determined by the variational technique of Brown and Collin using an artificial short. Their method is extended to each of the three types of discontinuity enumerated above. A cosine series of two terms has been used as a trial function. Formulas have been developed to simplify the usual Weissfloch calculation.

INTRODUCTION

AN ARRAY of spaced dielectric strips constitutes an anisotropic medium on account of the difference in the propagation constants of waves polarized parallel to and perpendicular to the plane of the strips. The medium can therefore be used as a polarizer for the conversion of linear to circular polarization. It is expedient, in order to reduce reflections, to cut a step near the edge of the strips to approximate a quarter-wave transformer, as shown in Fig. 1(a).

Practical considerations require the use of a protective covering for a polarizer, therefore leading to a final surface structure as shown in Fig. 1(b). A simultaneous match by both polarizations at a single surface of the type in Fig. 1(a) is not physically realizable. It can be shown that a simultaneous match will be more closely approximated by the surface of Fig. 1(b), provided

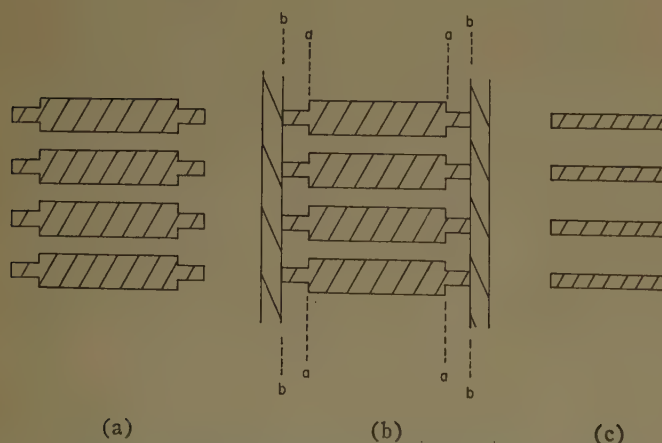


Fig. 1—(a) Cross section of dielectric strip polarizer. (b) Dielectric strip polarizer with skin showing discontinuities to *a-a* and *b-b*. (c) Dielectric strip-free space discontinuity.

† Radio and Elec. Engrg. Div., Natl. Res. Council, Ottawa, Can.

that a covering of the correct thickness is used. This result has some practical interest.

Therefore, the boundary discontinuities at *a-a* and *b-b* of Fig. 1(b) pertain directly to the design of a polarizer, and form the subject of this paper. From elementary theory based on the respective propagation constants of the two regions, the magnitude of the reflection and transmission coefficients is readily calculated while the phase shifts are either 0° or 180°. Any departure from these elementary values affects the polarizer performance; therefore an exact calculation is desirable. The discontinuities will be specified in terms of the scattering coefficients $S_{11} = \alpha \exp(j\theta)$, $S_{12} = \beta \exp(j\phi)$, $S_{22} = -\alpha \exp[j(2\phi - \theta)]$, and $\alpha^2 + \beta^2 = 1$; the last two relations follow from the unitary character of the scattering matrix. The method employed is the variational technique of Brown and Collin^{1,2} and Collin^{3,4} using an artificial short. These authors have dealt extensively with the boundary shown in Fig. 1(c). The method is here extended to the boundaries *a-a* and *b-b* of Fig. 1(b). In addition, formulas and methods are given which are intended to facilitate the accurate calculation of the scattering coefficients.

STRIP-STRIP DISCONTINUITY

For the dielectric step, shown at $z=0$ in Fig. 2, e and d , the thickness and separation of the strips, respectively, will vary discontinuously while $c=e+d$ remains unchanged. Symbols for quantities pertaining to the structure in the region $z < 0$ will be designated by a dash (') to distinguish them from quantities related to the region $z > 0$. The dimensions require that there be one propagating mode only in each region, with, however, propagation both toward and away from the discontinuity. This condition implies one imaginary root only, p_1 and p_1' , of (51) and is satisfied by,

$$e/\lambda\sqrt{\epsilon} < 1 \quad (1)$$

where ϵ is the relative dielectric constant of the strips, and λ is the free-space wavelength.

Starting with the formulation in terms of the electric

¹ R. E. Collin and J. Brown, "The calculation of the equivalent circuit of an axially unsymmetrical waveguide junction," Monograph No. 145R, *Proc. IEE*, pp. 121-128; August, 1955.

² R. E. Collin and J. Brown, "The design of quarter-wave matching layers for dielectric surfaces," Monograph No. 149R, *Proc. IEE*, pp. 153-158; September, 1955.

³ R. E. Collin, Ph.D. dissertation, University of London, Eng.; 1953.

⁴ R. E. Collin, "Reflection and transmission at a slotted dielectric interface," *Can. J. Phys.*, vol. 34, pp. 398-411; April, 1956.

field, (47), the magnetic field on each side of the discontinuity may be obtained from

$$\frac{\partial H_y}{\partial z} = -j\omega\epsilon E_x.$$

Manipulation of the equation thus obtained leads to a common type of equation,

$$\begin{aligned} \frac{Y}{p_1'} \left(\int F_1' \epsilon E_x dx \right)^2 - \sum_{m=2}^{\infty} \frac{1}{p_m'} \left(\int F_m' \epsilon E_x dx \right)^2 \\ = \frac{Y_s}{p_1} \left(\int F_1 \epsilon E_x dx \right)^2 + \sum_{m=2}^{\infty} \frac{1}{p_m} \left(\int F_m \epsilon E_x dx \right)^2. \quad (2) \end{aligned}$$

Y , the admittance of the propagating mode, is defined by

$$Y = \frac{A_1' - B_1'}{A_1' + B_1'}$$

where a similar definition of Y_s . Y_s is subsequently made a pure imaginary according to the method of Collin and Brown. Eq. (2) is variational in the sense that a variation in E_x from its true value produces a second-order effect in Y .

Starting with the expansion in terms of the magnetic field, (50), a similar equation results.

$$\begin{aligned} Z p_1' \left(\int \frac{G_1' H_y dx}{\epsilon} \right)^2 - \sum_{m=2}^{\infty} p_m' \left(\int \frac{G_m' H_y dx}{\epsilon} \right)^2 \\ = Z_s p_1 \left(\int \frac{G_1 H_y dx}{\epsilon} \right)^2 + \sum_{m=2}^{\infty} p_m \left(\int \frac{G_m H_y dx}{\epsilon} \right)^2, \quad (3) \end{aligned}$$

where

$$Z = \frac{C_1' - D_1'}{C_1' + D_1'} \quad \text{and} \quad Z_s = \frac{C_1 - D_1}{C_1 + D_1}.$$

It will be seen that Y , Y_s , Z , Z_s , relate the electric and magnetic fields associated with the propagating mode.

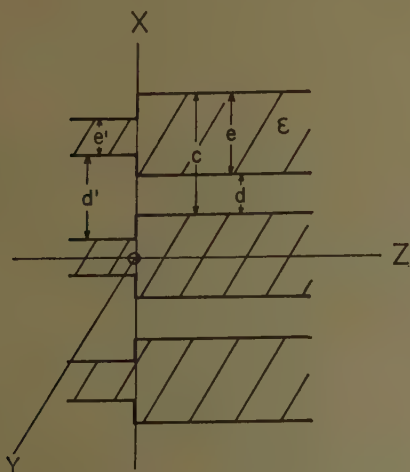


Fig. 2—The discontinuity at a step in the dielectric strips.

They may be interpreted and manipulated in the manner usual for an admittance or impedance. Hence (2) and (3) place upper and lower limits on the susceptance $B = -jY$.

The development for parallel polarization is carried through similarly by matching E_y and H_x across the discontinuity. The resulting expressions are

$$\begin{aligned} Y q_1 \left(\int F_1' E_y dx \right)^2 - \sum_{m=2}^{\infty} q_m' \left(\int F_m' E_y dx \right)^2 \\ = Y_s q_1 \left(\int F_1 E_y dx \right)^2 + \sum_{m=2}^{\infty} q_m \left(\int F_m E_y dx \right)^2 \quad (4) \end{aligned}$$

and

$$\begin{aligned} \frac{Z}{q_1'} \left(\int G_1' H_x dx \right)^2 - \sum_{m=2}^{\infty} \frac{1}{q_m'} \left(\int G_m' H_x dx \right)^2 \\ = \frac{Z_s}{q_1} \left(\int G_1 H_x dx \right)^2 + \sum_{m=2}^{\infty} \frac{1}{q_m} \left(\int G_m H_x dx \right)^2. \quad (5) \end{aligned}$$

STRIP-SKIN DISCONTINUITY

For perpendicular polarization, the fields in the strip region of Fig. 3 are given by (47), together with the corresponding equation for the magnetic field. The seare to be matched with the fields in the uniform dielectric of the skin.

$$\begin{aligned} E_x = a_0 e^{-ikz} + b_0 e^{ikz} \\ + \sum_{r=1}^{\infty} (a_r e^{-k_r z} + b_r e^{k_r z}) \cos \frac{2\pi r x}{c}, \quad (63) \end{aligned}$$

and as a consequence of $(\partial H_y)/(\partial z) = -j\omega\epsilon' E_x$,

$$\begin{aligned} H_y = j\omega\epsilon' \left[\frac{a_0 e^{-ikz} - b_0 e^{ikz}}{jk} \right. \\ \left. + \sum_{r=1}^{\infty} \frac{a_r e^{-k_r z} - b_r e^{k_r z}}{k_r} \cos \frac{2\pi r x}{c} \right]. \quad (6) \end{aligned}$$

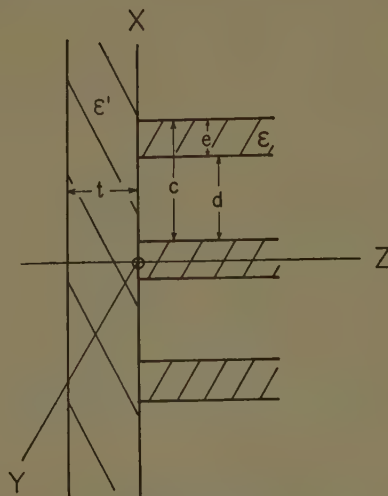


Fig. 3—The skin-strip discontinuity.

Matching the fields at $z=0$,

$$\epsilon' \left[\frac{a_0 - b_0}{jk} + \sum_{r=1}^{\infty} \frac{a_r - b_r}{k_r} \cos \frac{2\pi r x}{c} \right] = \frac{1}{k_0} \left[\frac{(A_1 - B\hat{\epsilon})\epsilon F_1}{p_1} + \sum_{m=2}^{\infty} \frac{A_m \hat{\epsilon} F_m}{p_m} \right]. \quad (7)$$

Multiplying (7) by $E_x dx$ and integrating, and making use of (65),

$$\begin{aligned} & \frac{-jkY'}{c} \left(\int E_x dx \right)^2 + \frac{2k^2}{c} \sum_{r=1}^{\infty} \frac{Y_r}{k_r} \left(\int E_x \cos \frac{2\pi r x}{c} dx \right)^2 \\ &= k_0 \left[\frac{Y_s}{p_1} \left(\int F_1 \hat{\epsilon} E_x dx \right)^2 + \sum_{m=2}^{\infty} \frac{1}{p_m} \left(\int F_m \hat{\epsilon} E_x dx \right)^2 \right]. \quad (8) \end{aligned}$$

Starting with expressions for the magnetic field,

$$\begin{aligned} & \frac{jZ'}{kc} \left(\int H_y dx \right)^2 + \frac{2}{k^2 c} \sum_{r=1}^{\infty} Z_r k_r \left(\int H_y \cos \frac{2\pi r x}{c} dx \right)^2 \\ &= \frac{1}{k_0} \left[p_1 Z_s \left(\int \frac{G_1 H_y dx}{\hat{\epsilon}} \right)^2 + \sum_{m=2}^{\infty} p_m \left(\int \frac{H_y G_m dx}{\hat{\epsilon}} \right)^2 \right]. \quad (9) \end{aligned}$$

For parallel polarization the corresponding equations are

$$\begin{aligned} & \frac{jkY'}{c} \left(\int E_y dx \right)^2 + \frac{2}{c} \sum_{r=1}^{\infty} k_r Y_r \left(\int E_y \cos \frac{2\pi r x}{c} dx \right)^2 \\ &= k_0 \left[q_1 Y_s \left(\int E_y F_1 dx \right)^2 + \sum_{m=0}^{\infty} q_m \left(\int E_y F_n dx \right)^2 \right] \quad (10) \end{aligned}$$

and

$$\begin{aligned} & \frac{-jZ'}{kc} \left(\int H_x dx \right)^2 + \frac{2}{c} \sum_{r=1}^{\infty} \frac{Z_r}{k_r} \left(\int H_x \cos \frac{2\pi r x}{c} dx \right)^2 \\ &= \frac{1}{k_0} \left[\frac{Z_s}{q_1} \left(\int G_1 H_x dx \right)^2 + \sum_{m=2}^{\infty} \frac{1}{q_m} \left(\int G_m H_x dx \right)^2 \right]. \quad (11) \end{aligned}$$

Eqs. (8)–(11) differ from those previously considered by Brown and Collin only in the presence of the admittance factors Y_r . These quantities are directly calculable on the grounds that, while the higher modes may travel in both directions within the skin, in free space adjacent to the skin they travel only in a direction away from the skin. The admittance just inside the skin adjacent to free space is, for perpendicular polarization,

$$Y_{0r} = -\frac{k_r}{\epsilon' k_{0r}},$$

where

$$k_{0r}^2 = 4\pi^2 \left(\frac{r^2}{c^2} - \frac{1}{\lambda^2} \right);$$

and therefore at $z=0$, assuming that there is a skin thickness t .

$$Y_r = \frac{-\frac{k_r}{\epsilon' k_{0r}} - \tanh k_r t}{1 + \frac{k_r}{\epsilon' k_{0r}} \tanh k_r t} \quad (12)$$

and

$$Z_r = \frac{1}{Y_r}.$$

For parallel polarization, $Y_{0r} = -k_{0r}/k_r$, and at $z=0$,

$$Y_r = \frac{-\frac{k_{0r}}{k_r} - \tanh k_r t}{1 + \frac{k_{0r}}{k_r} \tanh k_r t} \quad (13)$$

and

$$Z_r = \frac{1}{Y_r}.$$

For thicknesses such that $k_r t \ll 1$, approximate formulas are applicable as follows.

For perpendicular polarization,

$$Y_r \approx -\frac{k_r}{\epsilon' k_{0r}} - \frac{k_r t (\epsilon' - 1)}{\epsilon'^2 (h_r^2 - 1)} [(\epsilon' + 1) h_r^2 - \epsilon'] \quad (14)$$

$$Z_r \approx -\frac{\epsilon' k_{0r}}{k_r} + \frac{k_0^2 t (\epsilon' - 1)}{k_r} [(\epsilon' + 1) h_r^2 - \epsilon'], \quad (15)$$

where $h_r = r\lambda/c$.

For parallel polarization,

$$Y_r \approx -\frac{k_{0r}}{k_r} + \frac{k_0^2 t}{k_r} (\epsilon' - 1) \quad (16)$$

$$Z_r \approx -\frac{k_r}{k_{0r}} - \frac{k_r t (\epsilon' - 1)}{h_r^2 - 1}. \quad (17)$$

The results of the variational calculations yield Y' and Z' which are the admittance and impedance functions within the skin. It is the corresponding functions outside the skin which are of interest in determining the scattering coefficients of the discontinuity. If the external admittance Y which is required is referred to the plane $z=0$, then Y' and Y may be referred to the plane $z=-t$ where they can be related in the ratio k/k_0 across the dielectric-space boundary, giving

$$k_0 \frac{Y + j \tan k_0 t}{1 + j Y \tan k_0 t} = k \frac{Y' + j \tan k t}{1 + j Y' \tan k t}. \quad (18)$$

For a thin skin, (18) reduces to

$$Y \approx \frac{k}{k_0} Y' + jk_0 t(\epsilon' - 1). \quad (19)$$

Eqs. (18) or (19) may be used to complete the boundary solution. However, it is more convenient to consider the skin separately. If the skin were separated to a great distance from the strips, an admittance Y' just within the right-hand side would be equivalent to an admittance Y_0 just outside, where

$$Y_0 = \frac{k}{k_0} Y'. \quad (20)$$

The procedure is represented by the equivalent structure, Fig. 4, where Y_0 is the admittance in the infinitesimal separation between the skin and the strips. The wave amplitudes equivalent to Y can then be related to those equivalent to Y_0 by transmission line relationships. If

$$\begin{pmatrix} A \\ B \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} A_0 \\ B_0 \end{pmatrix}$$

are the respective amplitudes referred to the right-hand face of the skin,

$$\begin{pmatrix} A \\ B \end{pmatrix} = (M) \begin{pmatrix} A_0 \\ B_0 \end{pmatrix}, \quad (21)$$

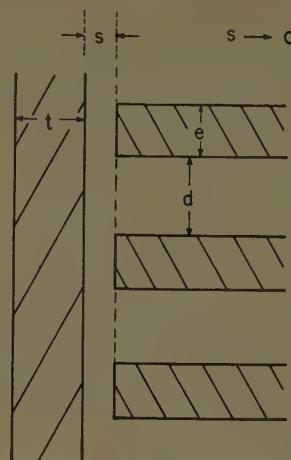


Fig. 4—An equivalent structure for the skin-strip discontinuity.

$$1 + \sum_{n=1}^N \alpha_n \cos \frac{2\pi nx}{c}$$

and the magnetic field components,

$$1 + \sum_{n=1}^N \beta_n \cos \frac{2\pi nx}{c}.$$

For perpendicular polarization the following terms, the values of which are written down without derivation, are generated.

$$A_{mn} = \int F_m \cos \frac{2\pi nx}{c} \hat{x} dx = \frac{(\epsilon - 1) \left\{ \frac{h_n \lambda \epsilon}{\pi} \sin \frac{n\pi \epsilon}{c} + e(\epsilon + p_m^2)(p_m^2 - h_n^2) T_m \cos \frac{n\pi e}{c} \right\}}{\epsilon^{1/2} (p_m^2 + \epsilon - h_n^2)(p_m^2 + 1 - h_n^2) \left[\frac{e}{H(\beta_m d e)} + \frac{\epsilon d}{H(\beta_m d)} \right]^{1/2}}, \quad (23)$$

where

$$(M) = \begin{bmatrix} [\cos kt + \frac{1}{2}j(x + \frac{1}{x}) \sin kt] e^{-jk_0 t} & \frac{1}{2}j(x - \frac{1}{x}) \sin k t e^{-jk_0 t} \\ -\frac{1}{2}j(x - \frac{1}{x}) \sin k t e^{jk_0 t} & [\cos kt - \frac{1}{2}j(x + \frac{1}{x}) \sin kt] e^{jk_0 t} \end{bmatrix} \quad (22)$$

and

$$x = \sqrt{\epsilon'}.$$

THE TRIAL FUNCTION

Cosine series have been used throughout as trial functions. Thus, the electric field components have been written,

where

$$T_m = \frac{\tan \frac{\beta_m d e}{2}}{\frac{\beta_m d e}{2}},$$

$$H(u) = \frac{1 + \cos u}{1 + \frac{\sin u}{u}}$$

and

$$C_{mn} = \int G_m \cos \frac{2\pi nx}{c} \frac{1}{\hat{e}} \cdot dx = \frac{\epsilon - 1}{\epsilon^{1/2}} \frac{\left\{ \frac{h_n \lambda}{\pi} (p_m^2 + \epsilon + 1 - h_n^2) \sin \frac{n\pi \epsilon}{c} - e(\epsilon + p_m^2) T_m \cos \frac{n\pi e}{c} \right\}}{(p_m^2 + \epsilon - h_n^2)(p_m^2 + 1 - h_n^2) \left[\frac{e}{H(\beta_m d e)} + \frac{\epsilon d}{H(\beta_m d)} \right]^{1/2}}. \quad (24)$$

For parallel polarization, the corresponding terms are

$$A_{mn} = \int F_m \cos \frac{2\pi nx}{c} dx = \frac{(\epsilon - 1) \left\{ \frac{h_n \lambda}{c} \sin \frac{n\pi e}{c} - e(\epsilon + q_m^2) T_m \cos \frac{n\pi e}{c} \right\}}{(q_m^2 + \epsilon - h_n^2)(q_m^2 + 1 - h_n^2) \left[\frac{e}{H(\beta_m d e)} + \frac{d}{H(\beta_m 0 d)} \right]^{1/2}} \quad (25)$$

and

$$C_{mn} = \int G_m \cos \frac{2\pi nx}{c} dx = A_{mn}. \quad (26)$$

There occur also in the skin-strip formulas, integrals of the form,

$$\begin{aligned} \int \cos \frac{2\pi rx}{c} \cos \frac{2\pi nx}{c} dx &= 0 \quad (\text{for } r \neq n) \\ &= c/2 \quad (\text{for } r = n \neq 0) \\ &= c \quad (\text{for } r = n = 0). \end{aligned}$$

Substituting in (8),

$$\begin{aligned} \frac{-jkY'}{c} c^2 + \frac{2k^2}{c} \sum_{n=1}^N \frac{Y_n}{k_n} \left(\alpha_n \frac{c}{2} \right)^2 \\ = k_0 \left[\frac{Y_s}{p_1} (A_{10} + \alpha_1 A_{11} + \dots)^2 \right. \\ \left. + \sum_{m=2}^{\infty} \frac{1}{p_m} (A_{m0} + \alpha_1 A_{m1} + \dots)^2 \right]. \quad (27) \end{aligned}$$

Since Y' is variational, $(\partial Y')/(\partial \alpha_n) = 0$, which on substitution in (27) leads to the set of equations,

$$\begin{aligned} P_{0n} + \alpha_1 P_{1n} + \dots + \alpha_n \left[P_{nn} - \frac{c\epsilon' k_0 Y_n}{2k_n} \right] \\ + \dots + \alpha_N P_{Nn} = 0 \quad n = 1, 2, \dots, N, \quad (28) \end{aligned}$$

where

$$P_{ij} = \sum_{m=1}^{\infty} \frac{Y_{sm} A_{mi} A_{mj}}{p_m},$$

and

$$\begin{aligned} Y_{sm} &= Y_s \quad \text{for } m = 1, \\ Y_{sm} &= 1 \quad \text{for } m \neq 1. \end{aligned}$$

Eq. (27) can then be reduced using (28) and (14).

$$\begin{aligned} -jkY' = k_0 \left[P_{00} - \sum_{n=1}^N \alpha_n^2 \left\{ P_{nn} + \frac{c}{2\sqrt{h_n^2 - 1}} \right. \right. \\ \left. \left. + \frac{ck_0 l(\epsilon' - 1)}{2} \left(1 + \frac{h_n^2}{\epsilon'(h_n^2 - 1)} \right) \right\} \right]. \quad (29) \end{aligned}$$

Similarly, (9) leads to

$$\begin{aligned} \frac{jcZ'}{k} = \frac{1}{k_0} \left[Q_{00} - \sum_{n=1}^N \beta_n^2 \left\{ Q_{nn} + \frac{c\sqrt{h_n^2 - 1}}{2} \right. \right. \\ \left. \left. - \frac{ck_0 l(\epsilon' - 1)}{2} \left[h_n^2 \left(1 + \frac{1}{\epsilon'} \right) - 1 \right] \right\} \right] \quad (30) \end{aligned}$$

with

$$\begin{aligned} Q_{0n} + \beta_1 Q_{1n} + \dots + \beta_n \left[Q_{nn} - \frac{Z_n k_n c}{2\epsilon' k_0} \right] \\ + \dots + \beta_N Q_{Nn} = 0, \quad (31) \end{aligned}$$

where

$$Q_{ij} = \sum p_m Z_{sm} C_{mi} C_{mj}.$$

Eq. (10) leads to

$$\begin{aligned} jkcY' = k_0 \left[P_{00} - \sum_{n=1}^N \alpha_n^2 \left\{ P_{nn} + \frac{c}{2} \sqrt{h_n^2 - 1} \right. \right. \\ \left. \left. - \frac{c}{2} k_0 l(\epsilon' - 1) \right\} \right] \quad (32) \end{aligned}$$

with

$$\begin{aligned} P_{0n} + \alpha_1 P_{1n} + \dots + \alpha_n \left(P_{nn} - \frac{c}{2} \frac{k_n Y_n}{k_0} \right) \\ + \dots + \alpha_N P_{Nn} = 0, \quad (33) \end{aligned}$$

where

$$P_{ij} = \sum_{m=1}^{\infty} q_m Y_{ms} A_{mi} A_{mj}.$$

Eq. (11) leads to

$$\begin{aligned} \frac{-jcZ'}{k} = \frac{1}{k_0} \left[Q_{00} - \sum_{n=1}^N \beta_n^2 \left\{ Q_{nn} + \frac{c}{2\sqrt{h_n^2 - 1}} \right. \right. \\ \left. \left. + \frac{ck_0 l(\epsilon' - 1)}{2(h_n^2 - 1)} \right\} \right] \quad (34) \end{aligned}$$

with

$$\begin{aligned} Q_{0n} + \beta_1 Q_{1n} + \dots + \beta_n \left\{ Q_{nn} - \frac{ck_0 Z_n}{2k_n} \right\} \\ + \dots + \beta_N Q_{Nn} = 0, \quad (35) \end{aligned}$$

where

$$Q_{ij} = \sum_{m=1}^{\infty} \frac{Z_{sm} C_{mi} C_{mj}}{q_m}.$$

It will be observed that the admittance and impedance formulas (29), (30), (32), and (34) involve the skin parameters only in the terms containing $k_0 l(\epsilon' - 1)$. The inclusion of these terms requires a very small additional effort in the calculation.

The reduction of the formulas for the strip-strip discontinuity leads to the following results. For perpendicular polarization, (2) becomes

$$\begin{aligned} & \frac{Y}{p_1'} (A_{10}'^2 - \alpha_1^2 A_{11}'^2 - \dots) \\ & - \sum_{m=2}^{\infty} \frac{1}{p_m'} (A_{m0}'^2 - \alpha_1^2 A_{m1}'^2 - \dots) \\ & = \frac{Y_s}{p_1} (A_{10}^2 - \alpha_1^2 A_{11}^2 + \dots) \\ & + \sum_{m=2}^{\infty} \frac{1}{p_m} (A_{m0}^2 - \alpha_1^2 A_{m1}^2 - \dots), \quad (36) \end{aligned}$$

where

$$\begin{aligned} & \frac{Y}{p_1'} (A_{10}' + \alpha_1 A_{11}' + \dots) A_{1n}' \\ & - \sum_{m=2}^{\infty} \frac{1}{p_m'} (A_{m0}' + \alpha_1 A_{m1}' + \dots) A_{mn}' \\ & = \frac{Y_s}{p_1} (A_{10} + \alpha_1 A_{11} + \dots) A_{1n} \\ & + \sum_{m=2}^{\infty} \frac{1}{p_m} (A_{m0} + \alpha_1 A_{m1} + \dots) A_{mn} \\ & n = 1, 2, \dots, N. \quad (37) \end{aligned}$$

The corresponding expressions resulting from (3), (4), and (5) can be immediately written down. Unfortunately, (37) contains the unknown admittance Y . The laborious procedure of successive approximations is presumably necessary.

THE SCATTERING COEFFICIENTS

The problem of translating the admittances Y for various values of Y_s into the scattering coefficients of the discontinuity is common to all of the structures and polarizations. Brown and Collin have used multiple values of Y_s and Y and have plotted the position of the short vs the position of the minimum according to the Weissfloch procedure. In principle, three values of Y , if known accurately, are sufficient. However, the attempt to use only three approximate values has given very inaccurate results. The following procedure uses four approximate values and appears to give excellent results. However, it is not known how closely the method approaches the optimum use of the available information. Let B_s , where $Y_s = jB_s$, correspond to forward and backward traveling waves represented at the plane of the discontinuity by the matrix

$$\begin{pmatrix} A \\ B \end{pmatrix}.$$

Then

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} e^{jk_s} \\ -e^{-jk_s} \end{pmatrix}$$

where $ks = -\cot^{-1}B_s$. On the other side of the discontinuity, the susceptance B , where $Y = jB$, corresponds to

$$\begin{pmatrix} A' \\ B' \end{pmatrix} = \begin{pmatrix} ae^{-jk_d} \\ -ae^{jk_d} \end{pmatrix}$$

where $kd = \cot^{-1}B$, and a is an unspecified real constant. The two matrices are related by the scattering properties of the discontinuity,

$$\begin{bmatrix} ae^{-jk_d} \\ -ae^{jk_d} \end{bmatrix} = \frac{1}{\beta} \begin{bmatrix} e^{-j\phi} & \alpha e^{j(\phi-\theta)} \\ \alpha e^{j(\theta-\phi)} & e^{j\phi} \end{bmatrix} \begin{bmatrix} e^{jk_s} \\ -e^{-jk_s} \end{bmatrix},$$

from which,

$$ae^{jk_d} = \frac{1}{\beta} \{ e^{j(\phi-k_s)} - \alpha e^{j(\theta-\phi+k_s)} \}. \quad (38)$$

Two values of Y_s , Y_{s1} , and Y_{s2} are chosen so that

$$Y_{s1} \cdot Y_{s2} = 1. \quad (39)$$

In writing (83) for the two values of s , performing the necessary multiplications, and noting that $e^{j(k_{s1}-s_2)} = j$, it follows that

$$a_1 a_2 e^{jk(d_1+d_2)} = \frac{1}{\beta^2} e^{j(2\phi-k(s_1+s_2))} [1 + \alpha^2 e^{2j(\theta-2\phi+k(s_1+s_2))}] \quad (40)$$

and

$$\begin{aligned} & a_1/a_2 e^{jk(d_1-d_2)} \\ & = -\frac{1}{\beta^2} [j(1-\alpha^2) + 2\alpha \cos(\theta-2\phi+k(s_1+s_2))]. \quad (41) \end{aligned}$$

From (41),

$$\cot k(d_1-d_2) = \frac{2\alpha}{1-\alpha^2} \cos(\theta-2\phi+k(s_1+s_2)). \quad (42)$$

For a second pair of values,

$$\cot k(d_1'-d_2') = \frac{2\alpha}{1-\alpha^2} \cos(\theta-2\phi+k(s_1'+s_2')). \quad (43)$$

Eliminating α from (42) and (43),

$$\begin{aligned} & \tan(\theta-2\phi+k(s_1+s_2)) \\ & = \cot \Delta - \frac{\cot k(d_1'-d_2')}{\cot k(d_1-d_2)} \csc \Delta, \quad (44) \end{aligned}$$

where

$$\Delta = k(s_1' + s_2' - s_1 - s_2).$$

From (42),

$$\frac{2\alpha}{1-\alpha^2} = \frac{\cot k(d_1-d_2)}{\cos(\theta-2\phi+k(s_1+s_2))}, \quad (45)$$

and from (40),

$$2\phi = k(d_1 + d_2 + s_1 + s_2) + \delta, \quad \left. \begin{array}{l} \text{where} \\ \delta = \tan^{-1} \left\{ \frac{\alpha^2 \sin 2(\theta - 2\phi + k(s_1 + s_2))}{1 + \alpha^2 \cos 2(\theta - 2\phi + k(s_1 + s_2))} \right\} \\ = \alpha^2 \sin 2(\theta - 2\phi + k(s_1 + s_2)) + O(\alpha^4). \end{array} \right\} \quad (46)$$

There is a similar equation available for 2ϕ from the values d_1', d_2', s_1', s_2' . A comparison of the two is a useful measure of the consistency of the results. Eqs. (44), (45), and (46) then provide a set of simple exact equations for determining θ , ϕ , and α in terms of known quantities.

A CALCULATED EXAMPLE

The results of a calculated example are shown in Tables I and II. The trial function in each case included two terms only. Departures from the elementary solutions are rather slight as previously pointed out by Collin. However, θ must be taken into account to obtain the highest degree of cancellation for a reflectionless surface. The ϕ 's, if cumulative, would require a significant correction in polarizer design, but in examples calculated to date they have tended to cancel. There is also a significant departure from the elementary solution in α_{\perp} for the strip-strip discontinuity.

TABLE I

| Skin-Strip | Strip-Strip |
|-------------------------------|----------------------|
| $e/\lambda = 0.088$ | $e'/\lambda = 0.088$ |
| $d/\lambda = 0.468$ | $d'/\lambda = 0.468$ |
| $t/\lambda = 0.012$ | $e/\lambda = 0.244$ |
| $\epsilon' = \epsilon = 2.68$ | $d/\lambda = 0.312$ |
| | $\epsilon = 2.68$ |

TABLE II

| Variational | Elementary | Variational | Elementary |
|-----------------------------|------------|------------------------------|------------|
| α_{\perp} 0.0289 | 0.0292 | α_{\perp} 0.0649 | 0.0750 |
| θ_{\perp} 186°40' | 180° | θ_{\perp} 189°39' | 180° |
| ϕ_{\perp} -19.2' | 0° | ϕ_{\perp} 20' | 0° |
| α_{\parallel} 0.0704 | 0.0700 | α_{\parallel} 0.0911 | 0.0915 |
| θ_{\parallel} 173°3' | 180° | θ_{\parallel} 176°37' | 180° |
| ϕ_{\parallel} 5.7' | 0° | ϕ_{\parallel} 1.4' | 0° |

It has been noted by Angulo⁵ that a variational solution such as (11) using a constant trial function leads to an equivalent circuit consisting of a transformer plus a shunt susceptance. Similarly, (3) would lead to a transformer plus a series reactance. These circuits imply the opposite correlation between the reflection and transmission coefficients at the boundary. It seems reasonable, therefore, to assume that a constant trial function will give no information regarding the transmission coefficient. This conclusion is borne out by experience in calculating the terms.

⁵ C. M. Angulo, "Discontinuities in a rectangular waveguide partially filled with dielectric," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-5, pp. 68-74; January, 1957.

APPENDIX

MODES IN THE STRIP STRUCTURE

Perpendicular Polarization

Referred to the coordinate system of Fig. 5, the electric fields in the strip structure can be written

$$E_x = \sum_{m=1}^{\infty} (A_m e^{-r_m z} + B_m e^{+r_m z}) F_m(x), \quad (47)$$

where the functions $F_m(x)$ satisfy the equations,

$$\frac{d^2 F_m}{dx^2} + \beta_{md}^2 F_m = 0; \quad \beta_{md}^2 = k_0^2 \epsilon + r_m^2 = k_0^2 (\epsilon + p_m^2), \quad (48)$$

and

$$\frac{d^2 F_m}{dx^2} + \beta_{m0}^2 F_m = 0; \quad \beta_{m0}^2 = k_0^2 + r_m^2 = k_0^2 (1 + p_m^2). \quad (49)$$

Eq. (48) refers to the dielectric region and (49) to the free space region, k_0 being the free space propagation constant. The problem may similarly be set up in terms of the magnetic field,

$$H_y = \sum_{m=1}^{\infty} (C_m e^{-r_m z} + D_m e^{+r_m z}) G_m(x). \quad (50)$$

The analysis, which is conventional, is continued by matching the fields across the dielectric-free space boundary. There results a transcendental equation for the propagation constants,

$$\beta_{md} \tan \frac{\beta_{md} e}{0} = -\epsilon \beta_{m0} \tan \frac{\beta_{m0} d}{2}. \quad (51)$$

The orthonormal property of the functions $F(x)$ and $G(x)$ is readily established as

$$\left. \begin{array}{l} \int_{-c/2}^{c/2} \hat{\epsilon} F_m F_n dx = \delta_{mn} \\ \int_{-c/2}^{c/2} \frac{G_m G_n}{\hat{\epsilon}} dx = \delta_{mn} \end{array} \right\}, \quad (52)$$

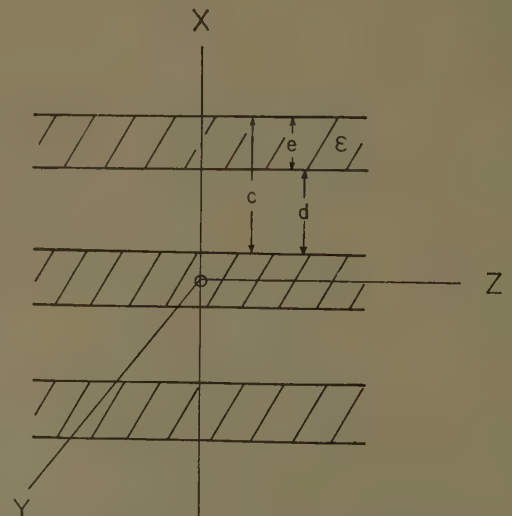


Fig. 5—Parallel dielectric strips with the coordinate system.

where $\delta_{mn}=0$ if $m \neq n$, and $\delta_{mn}=1$ if $m=n$; $\hat{\epsilon} \equiv \epsilon(x)$ where $\epsilon(x)=\epsilon$ in the dielectric and $\epsilon(x)=1$ in free space. It follows that in terms of the fields in the plane $z=0$,

$$A_m + B_m = \int \hat{\epsilon} E_x F_m dx \quad (53)$$

and

$$C_m + D_m = \int \frac{H_y G_m}{\hat{\epsilon}} dx. \quad (54)$$

Parallel Polarization

The formulation of the problem in the other polarization is similar but separate. In the summary which follows, many symbols already used are employed but no connection is implied between the two sets of formulas.

$$E_y = \sum_{m=1}^{\infty} (A_m e^{-s_m z} + B_m e^{+s_m z}) F_m(x) \quad (55)$$

where, in the dielectric,

$$\frac{d^2 F_m}{dx^2} + \beta_{md} F_m = 0, \quad \beta_{md}^2 = k_0^2 \epsilon + s_m^2 = k_0^2 (\epsilon + q_m^2) \quad (56)$$

and in free space,

$$\frac{d^2 F_m}{dx^2} + \beta_{m0} F_m = 0, \quad \beta_{m0}^2 = k_0^2 + s_m^2 = k_0^2 (1 + q_m^2) \quad (57)$$

$$H_x = \sum_{m=1}^{\infty} (C_m e^{-s_m z} + D_m e^{+s_m z}) G_m(x) \quad (58)$$

$$\beta_{md} \tan \frac{\beta_{md} e}{2} = -\beta_{m0} \tan \frac{\beta_{m0} d}{2} \quad (59)$$

$$\left. \begin{aligned} \int_{-c/2}^{c/2} F_m F_n dx &= \delta_{mn} \\ \int_{-c/2}^{c/2} G_m G_n dx &= \delta_{mn} \end{aligned} \right\} \quad (60)$$

$$A_m + B_m = \int E_y F_m dx \quad (61)$$

and

$$C_m + D_m = \int H_x G_m dx. \quad (62)$$

MODES IN UNIFORM DIELECTRIC

Perpendicular Polarization

The field can be expanded in the Fourier series,

$$\begin{aligned} E_x &= a_0 e^{-ikz} + b_0 e^{ikz} \\ &+ \sum_{r=1}^{\infty} (a_r e^{-k_r z} + b_r e^{+k_r z}) \cos \frac{2\pi r x}{c} \end{aligned} \quad (63)$$

where

$$k^2 = \epsilon' k_0^2$$

and

$$k_r^2 = 4\pi^2 \left(\frac{r^2}{c^2} - \frac{\epsilon'}{\lambda^2} \right). \quad (64)$$

The mode amplitudes can be expressed in terms of the field at $z=0$,

$$\left. \begin{aligned} a_0 + b_0 &= \frac{1}{c} \int_{-c/2}^{c/2} E_x dx \\ a_r + b_r &= \frac{2}{c} \int_{-c/2}^{c/2} E_x \cos \frac{2\pi r x}{c} dx \end{aligned} \right\} \quad (65)$$

Parallel Polarization

Expressions for E_y have the same form as those for E_x .

On the Propagation of Electromagnetic Waves Through Anisotropic Layers

G. TYRAS† AND G. HELD†

Summary—The problem of a stratified lossy plasma subjected to a steady magnetic field is considered. The plasma consists of (m) homogeneous layers with arbitrary thicknesses. An arbitrarily polarized plane wave originating in the free space is incident normally on the plasma. Expressions are derived relating the components of the reflected wave and the transmitted wave to the incident wave. The analyses are carried out for the cases when there is no coupling between the ordinary and the extraordinary waves at the interfaces, i.e., when the steady magnetic field is either in the plane of the interface or perpendicular to it. The derived results are written in form suitable for digital machine calculations.

INTRODUCTION

THE problem of electromagnetic wave propagation through homogeneous anisotropic ionized media has been studied by many workers.¹ Appleton² and later Hartree³ derived the expression for the complex index of refraction, which is now known as the Appleton-Hartree formula.⁴ Taylor⁵ and Goubau⁶ published graphs showing the dependence of the complex index of refraction and polarization on the electron density and collision frequency.

In considering an inhomogeneous anisotropic plasma, some simplified profiles of electron concentration as a function of distance are usually assumed. Rydbeck⁷ investigated the transmission properties of a parabolic layer and obtained solutions for two special cases of propagation—along and at right angles to the magnetic field. Wilkes⁸ assumed a triangular profile and solved the problem of reflection of very long waves when the magnetic field is vertical. Stanley⁹ considered a two-

rectangular slab profile of electron density variation. Budden¹⁰ described two methods of obtaining numerical solutions of the differential equations which govern the reflection of long and very long radio waves from the ionosphere at vertical and oblique incidence.

In this paper we treat the inhomogeneous plasma as one consisting of a number of homogeneous layers of arbitrary thicknesses. No particular pattern for the variation in the electron density and the collision frequency is assumed and the results are left in a general form.

Even though the mathematically elegant solutions of simple analytic profiles of electron concentration are helpful in many cases, there are instances where they cannot approximate the actual conditions well enough. Stanley⁹ outlined some of the inadequacies of the parabolic and the exponential profiles in describing the properties of the ionosphere, and showed how a two-slab model, in spite of its apparent crudity, alleviates some of the difficulties. With the solution to the multi-layer problem available, one can approximate any electron density and collision frequency profiles with a desired degree of accuracy. Numerical computations of this sort are laborious and probably impractical without the help of a modern digital computer. However, once programmed, quick answers can be obtained to specific problems. (It takes less than one minute for type 704 IBM digital computer to run through one set of parameters for a problem consisting of twenty-four layers.)

PLANE WAVES AND SOME PLANE BOUNDARY VALUE PROBLEMS

The plane wave solution for an unbounded anisotropic plasma is well known.¹¹ Here we shall review it only very briefly for reference.

One has at his disposal the two Maxwell's curl equations:

$$\begin{aligned}\nabla \times \mathbf{E} &= -j\omega\mu_0\mathbf{H} \\ \nabla \times \mathbf{H} &= \mathbf{J}_T\end{aligned}\quad (1)$$

and the equation of motion of an electron

$$m\mathbf{u} + m\nu\mathbf{u} = e(\mathbf{E}e^{j\omega t} + \mathbf{u} \times \mathbf{B}_0). \quad (2)$$

† Aero-Space Div., Boeing Airplane Co., Seattle, Wash.

¹ An excellent account of the past work done in this area may be found in S. K. Mitra, "The Upper Atmosphere," The Asiatic Society, Calcutta, India, 2nd ed., pp. 185-219.

² E. V. Appleton, "Geophysical influences on the transmission of wireless waves," *Proc. Phys. Soc.*, vol. 37, pp. 16D-22D; 1925.

³ D. R. Hartree, "The propagation of electromagnetic waves in a stratified medium," *Proc. Cambridge Phil. Soc.*, vol. 25, pp. 97-120; 1929.

⁴ S. K. Mitra, *op. cit.*, p. 187.

⁵ M. Taylor, "The Appleton-Hartree formula and dispersion curves for the propagation of electromagnetic waves through an ionized medium in the presence of an external magnetic field," *Proc. Phys. Soc.*, Pt. I, vol. 45, pp. 245-265; March, 1933; Pt. II, vol. 46, pp. 408-435; May, 1934.

⁶ G. Goubau, "Zur Dispersionstheorie der Ionosphäre," *Hochfrequenztech. u. Elektroakust.*, vol. 45, p. 179; 1935.

⁷ O. E. H. Rydbeck, "On the propagation of radio waves," *Trans. Chalm. University of Tech.*, Gothenburg, Sweden, no. 74, 1948.

⁸ M. V. Wilkes, "The theory of reflexion of very long wireless waves from the ionosphere," *Proc. Roy. Soc. (London)*, ser. A., vol. 175, pp. 143-163, April, 1940; Also, "The oblique reflexion of very long wireless waves from the ionosphere," *Proc. Roy. Soc. (London)*, ser. A., vol. 189, pp. 130-146, March, 1947.

⁹ J. P. Stanley, "Ionospheric reflection of very long radio waves," *Can. J. Res.*, vol. 28, pp. 549-557; November, 1950.

¹⁰ K. G. Budden, "The numerical solution of differential equations governing reflexion of long radio waves from the ionosphere," *Proc. Roy. Soc. (London)*, ser. A., vol. 227, pp. 516-535; February, 1955.

¹¹ J. A. Ratcliffe, "The Magneto-Ionic Theory and Its Application to the Ionosphere," Cambridge University Press, Cambridge, Eng., pp. 15-20; 1959.

The connection between (1) and (2) is made via the expression for the total current in the medium

$$J_T = eN\mathbf{u} + j\omega\epsilon_0\mathbf{E} = j\omega\epsilon_0(\epsilon_{ij})\mathbf{E} \quad (3)$$

where (ϵ_{ij}) is the permittivity tensor. In the Cartesian coordinate system, with the steady magnetic field H_0 oriented as shown in Fig. 1, the permittivity tensor has

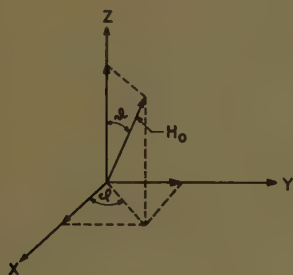


Fig. 1.

the form

$$(\epsilon_{ij}) = \begin{pmatrix} \epsilon + (\zeta - \epsilon) \sin^2 \theta \cos^2 \phi & \left(\frac{\zeta - \epsilon}{2}\right) \sin^2 \theta \sin 2\phi - j\eta \cos \theta & \left(\frac{\zeta - \epsilon}{2}\right) \sin 2\theta \cos \phi + j\eta \sin \theta \sin \phi \\ \left(\frac{\zeta - \epsilon}{2}\right) \sin^2 \theta \sin 2\phi + j\eta \cos \theta & \epsilon + (\zeta - \epsilon) \sin^2 \theta \sin^2 \phi & \left(\frac{\zeta - \epsilon}{2}\right) \sin 2\theta \sin \phi - j\eta \sin \theta \cos \phi \\ \left(\frac{\zeta - \epsilon}{2}\right) \sin 2\theta \cos \phi - j\eta \sin \theta \sin \phi & \left(\frac{\zeta - \epsilon}{2}\right) \sin 2\theta \sin \phi + j\eta \sin \theta \cos \phi & \zeta - (\zeta - \epsilon) \sin^2 \theta \end{pmatrix} \quad (4)$$

where ϵ , ζ , and η are all complex

$$\epsilon = \epsilon' - j\epsilon'', \quad \zeta = \zeta' - j\zeta'', \quad \eta = \eta' - j\eta'' \quad (5)$$

and

$$\left. \begin{aligned} \epsilon' &= 1 - \frac{p^2(1 - \sigma^2 + q^2)}{[(\sigma + 1)^2 + q^2][(\sigma - 1)^2 + q^2]} \\ \epsilon'' &= \frac{qp^2(1 + \sigma^2 + q^2)}{[(\sigma + 1)^2 + q^2][(\sigma - 1)^2 + q^2]} \\ \zeta' &= 1 - \frac{p^2}{1 + q^2}; \quad \zeta'' = \frac{qp^2}{1 + q^2} \\ \eta' &= \frac{\sigma p^2(\sigma^2 - 1 + q^2)}{[(\sigma + 1)^2 + q^2][(\sigma - 1)^2 + q^2]} \\ \eta'' &= \frac{2\sigma qp^2}{[(\sigma + 1)^2 + q^2][(\sigma - 1)^2 + q^2]} \end{aligned} \right\} \quad (6a)$$

The p 's and q 's and σ 's are given by

$$p = \frac{e}{\omega} \sqrt{\frac{N}{m\epsilon_0}}; \quad q = \frac{\nu}{\omega}; \quad \sigma = \frac{eB_0}{\omega m} \quad (6b)$$

One next solves the wave equation

$$\{\nabla \times \nabla \times - k_0^2(\epsilon_{ij})\} \mathbf{E} = 0 \quad (7)$$

for the case when the fields vary along one coordinate axis only, say z . In the case when the steady magnetic field is in the x - z plane, making an angle θ with the z -axis, one obtains for the electric field components

$$\begin{aligned} E_{x\pm} &= C_{o\pm} \exp(\pm jk_0 n^o z) + C_{e\pm} \exp(\pm jk_0 n^e z) \\ E_{y\pm} &= j[-C_{o\pm} \exp(-\alpha \pm jk_0 n^o z) + C_{e\pm} \exp(\alpha \pm jk_0 n^e z)] \\ E_{z\pm} &= \tan \theta \left[\left(\frac{n^{o2}}{\zeta} - 1 \right) C_{o\pm} \exp(\pm jk_0 n^o z) \right. \\ &\quad \left. + \left(\frac{n^{e2}}{\zeta} - 1 \right) C_{e\pm} \exp(\pm jk_0 n^e z) \right] \end{aligned} \quad (8)$$

where

$$\left. \frac{n^{o2}}{n^{e2}} \right\} = \frac{\epsilon \zeta}{\zeta \cos^2 \theta + \epsilon \sin^2 \theta} \left(1 \mp \frac{\eta}{\epsilon} \cos \theta e^{\mp \alpha} \right) \quad (9)$$

is the equivalent to the Appleton-Hartree formula,¹¹ and

$$\alpha = \ln \left[\left(\frac{\epsilon^2 - \eta^2 - \epsilon \zeta}{2\eta \zeta} \right) \frac{\sin^2 \theta}{\cos \theta} + \sqrt{1 + \left(\frac{\epsilon^2 - \eta^2 - \epsilon \zeta}{2\eta \zeta} \right)^2 \left(\frac{\sin^2 \theta}{\cos \theta} \right)^2} \right] \quad (10)$$

In the case when the steady magnetic field is in the x - y plane, making an angle ϕ with the x -axis, one obtains

$$\begin{aligned} E_{x\pm} &= C_{o\pm} \exp(\pm jk_0 n^o z) + C_{e\pm} \exp(\pm jk_0 n^e z) \\ E_{y\pm} &= C_{o\pm} \tan \phi \exp(\pm jk_0 n^o z) - C_{e\pm} \cot \phi \exp(\pm jk_0 n^e z) \\ E_{z\pm} &= j \frac{\eta}{\epsilon} \frac{C_{e\pm}}{\sin \phi} \exp(\pm jk_0 n^e z). \end{aligned} \quad (11)$$

Next, consider a plane air-plasma interface and a plane wave incident normally from the air half-space.

Let us denote the air as (0) region and plasma as (1) region and define the reflection and transmission coefficients R 's and T 's for the ordinary and the extraordinary waves as follows:

$$E_o^r = (R_o^{o1} + R_o^{e1}) E_o^i(0) \exp(jk_o z)$$

$$= (E_{oo}^r + E_{oe}^r) \exp(jk_o z)$$

$$E_{1\perp}^i = [T_o^{o1} \exp(-jk_1^o z) + T_e^{o1} \exp(-jk_1^e z)] E_o^i(0)$$

$$= E_{1o\perp}^i \exp(-jk_1^o z) + E_{1e\perp}^i \exp(-jk_1^e z) \quad (12)$$

where $k_1^o = k_o n_1^o$, $k_1^e = k_o n_1^e$ and \perp denotes the transverse field components only. Solving the boundary value problem, one finds the following expressions for R 's and T 's when the steady magnetic field is in the x - z plane:

$$R_o^{o1} = \frac{-R_o^{1o}}{2 \cosh \alpha_1} \begin{pmatrix} e^{\alpha_1} & j \\ -j & e^{-\alpha_1} \end{pmatrix} \quad (13)$$

$$R_e^{o1} = \frac{-R_o^{1o}}{2 \cosh \alpha_1} \begin{pmatrix} e^{-\alpha_1} & -j \\ j & e^{\alpha_1} \end{pmatrix} \quad (14)$$

$$T_{o,e}^{o1} = \left(1 - \frac{1}{R_{o,e}^{1o}}\right) R_{o,e}^{o1} \quad (15)$$

where

$$R_{o,e}^{1o} = \frac{n_1^{o,e} - 1}{n_1^{o,e} + 1} \quad (16)$$

The electric field in the plasma region is not purely transverse. Its longitudinal component can be found to be

$$E_{1z}^i = \frac{\tan \theta}{2 \cosh \alpha_1} \left\{ (1 - R_o^{1o}) \left(\frac{n_1^{o2}}{\xi_1} - 1 \right) (e^{\alpha_1} E_{ox}^i + j E_{oy}^i) \right. \\ \left. \cdot \exp(-jk_1^o z) \right. \\ \left. + (1 - R_e^{1o}) \left(\frac{n_1^{e2}}{\xi_1} - 1 \right) (e^{-\alpha_1} E_{ox}^i - j E_{oy}^i) \right. \\ \left. \cdot \exp(-jk_1^e z) \right\} \quad (17)$$

When the steady magnetic field is in the x - y plane, the corresponding expressions are

$$R_o^{o1} = -R_o^{1o} \begin{pmatrix} \cos^2 \phi & \sin \phi \cos \phi \\ \sin \phi \cos \phi & \sin^2 \phi \end{pmatrix} \quad (18)$$

$$R_e^{o1} = -R_o^{1o} \begin{pmatrix} \sin^2 \phi & -\sin \phi \cos \phi \\ -\sin \phi \cos \phi & \cos^2 \phi \end{pmatrix} \quad (19)$$

and the relations between T_o^{o1} and R_o^{o1} , and T_e^{o1} and R_e^{o1} are the same as before. The longitudinal component in this case is given by

$$E_{1z}^i = j \frac{\eta_1}{\epsilon_1} (1 - R_o^{1o}) (\sin \phi E_{ox}^i - \cos \phi E_{oy}^i) \\ \cdot \exp(-jk_1^e z). \quad (20)$$

In the alternate problem, when the wave originates in the plasma and is incident normally on an air space boundary, the ratio of the reflected to incident field components is simply equal to $R_{o,e}^{1o}$ for ordinary and the extraordinary waves, respectively. The ratio of the transmitted to the incident wave is equal to the transmission coefficient $T_{o,e}^{1o}$ given by

$$T_{o,e}^{1o} = 1 + R_{o,e}^{1o}. \quad (21)$$

The above relations hold for both cases of the orientation of the steady magnetic field, the x - z plane and x - y plane.

In the case of a plasma-plasma interface and the steady magnetic field in the x - y plane, the appropriate ratios are still of the simple form

$$R_{o,e}^{12} = \frac{n_1^{o,e} - n_2^{o,e}}{n_1^{o,e} + n_2^{o,e}} \\ T_{o,e}^{12} = 1 + R_{o,e}^{12} \quad (22)$$

for the ordinary and the extraordinary waves, respectively. When the steady magnetic field is in the x - z plane, the situation is much more complicated. In this case one finds that the reflection and transmission coefficients of (22) do not, in general, properly relate the reflected and the transmitted fields with the incident fields. One finds that the amplitudes of the various field components are related as follows:

$$\begin{pmatrix} C_{1o}^r \\ C_{1e}^r \end{pmatrix} = \begin{pmatrix} a_{oe} & -b_{eo} e^{\alpha_1} \\ b_{oe} e^{-\alpha_1} & a_{eo} \end{pmatrix} \begin{pmatrix} C_{1o}^i \\ C_{1e}^i \end{pmatrix} \quad (23)$$

and

$$\begin{pmatrix} C_{2o}^i \\ C_{2e}^i \end{pmatrix} = \begin{pmatrix} c_{oe} \exp\left(\frac{-\alpha_1 + \alpha_2}{2}\right) & -d_{eo} \exp\left(\frac{\alpha_1 + \alpha_2}{2}\right) \\ d_{oe} \exp\left(\frac{-\alpha_1 - \alpha_2}{2}\right) & c_{eo} \exp\left(\frac{\alpha_1 - \alpha_2}{2}\right) \end{pmatrix} \begin{pmatrix} C_{1o}^i \\ C_{1e}^i \end{pmatrix} \quad (24)$$

where

$$a_{oe} = \frac{R_o^{12} \cosh^2 \left(\frac{\alpha_1 + \alpha_2}{2} \right) + R_{oe}^{12} (T_e^{12} - R_{eo}^{12}) \sinh^2 \left(\frac{\alpha_1 - \alpha_2}{2} \right)}{\Delta}$$

$$b_{eo} = \frac{T_e^{12} (R_o^{12} - R_{oe}^{12}) \sinh \left(\frac{\alpha_1 - \alpha_2}{2} \right) \cosh \left(\frac{\alpha_1 + \alpha_2}{2} \right)}{\Delta} \quad (25)$$

$$c_{oe} = \frac{T_o^{12} \cosh \alpha_1 \cosh \left(\frac{\alpha_1 + \alpha_2}{2} \right)}{\Delta}$$

$$d_{eo} = \frac{T_e^{12} \cosh \alpha_1 \sinh \left(\frac{\alpha_1 - \alpha_2}{2} \right)}{\Delta} \quad (26)$$

$$\Delta = \cosh^2 \left(\frac{\alpha_1 + \alpha_2}{2} \right) + (T_o^{12} - R_{oe}^{12})(T_e^{12} - R_{eo}^{12}) \sinh^2 \left(\frac{\alpha_1 - \alpha_2}{2} \right) \quad (27)$$

and

$$R_{oe}^{12} = \frac{n_1^o - n_2^e}{n_1^o + n_2^o} \quad (28)$$

When the steady magnetic field is along the z -axis, the appropriate ratios of the field components may again be expressed in the form of (22).

THE STRATIFIED MEDIUM

Finally, consider a stratified anisotropic medium consisting of (m) finite-thickness homogeneous layers and backed up by an $(m+1)$ medium of infinite extent. The situation is depicted in Fig. 2. In the analysis that fol-

$$E_{m-i,\perp}^r(d_{m-i-1}) = \tanh \beta_{m-i} E_{m-i,\perp}^i(d_{m-i-1}) \quad (29)$$

$$E_{m-i,\perp}^i(d_{m-i-1}) = \left(\frac{1 + \tanh \gamma_{m-i-1}}{1 + \tanh \beta_{m-i} \tanh \gamma_{m-i-1}} \right) E_{m-i-1,\perp}^i(d_{m-i-1}) \quad (30)$$

where

$$\tanh \gamma_{m-i} = R^{m-i,m-i+1}$$

$$\tanh \beta_{m-i} = \tanh (\gamma_{m-i} + \beta_{m-i+1}) \exp (-j2\psi_{m-i}) \quad (31)$$

$$\psi_{m-i} = k_0 n_{m-i} l_{m-i}.$$

In region (1), $i=m-1$, we have to exercise more caution since at the interface $z=0$, the various field components are related by means of matrices rather than scalars. However, one can still establish for the ordinary wave, for instance, the following relations:

$$E_{10\perp}^i = \left(\frac{T_o^{o1}}{1 + \tanh \gamma_o^o \tanh \beta_1^o} \right) E_o^i(0) \quad (32)$$

$$E_{oo}^r = \left(R_o^{o1} + \frac{T_o^{1o} T_o^{o1}}{1 + \tanh \gamma_o^o \tanh \beta_1^o} \right) E_o^i(0), \quad (33)$$

where we have set $R_o^{1o} = -\tanh \gamma_o^o$. The corresponding expressions for the extraordinary wave are of identical form. The total reflected wave in region (0) is the sum of the contributions of the ordinary and the extraordinary waves which can now be combined; the result can be written in the form

$$E_o^r = (\rho_{ij}) \exp (jk_o z) E_o^i(0). \quad (34)$$

Next, we can write the components of (ρ_{ij}) more explicitly for the two cases of the transverse and the longitudinal magnetostatic field. In the transverse case, ($\theta = \pi/2$), one obtains

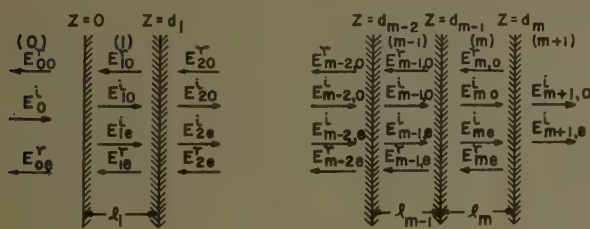


Fig. 2.

lows, we assume that there is no coupling between the ordinary and the extraordinary wave components at the interfaces. Thus, we consider only $\theta=0, \pi/2$ and ϕ arbitrary. Then the ordinary and the extraordinary components are simply related and we can use superposition throughout. Since the same analysis will apply to the ordinary and the extraordinary components, we can, for the time being, drop the subscripts o and e . One can show that for any i th layer, $0 \leq i < m-1$, the following relations between the various field components can be established:

$$\begin{aligned}\rho_{11} &= \tanh(\beta_1^o + \gamma_o^o) \cos^2 \phi + \tanh(\beta_1^e + \gamma_o^e) \sin^2 \phi \\ \rho_{12} &= \rho_{21} = [\tanh(\beta_1^o + \gamma_o^o) - \tanh(\beta_1^e + \gamma_o^e)] \sin \phi \cos \phi \\ \rho_{22} &= \tanh(\beta_1^o + \gamma_o^o) \sin^2 \phi + \tanh(\beta_1^e + \gamma_o^e) \cos^2 \phi. \quad (35)\end{aligned}$$

In the longitudinal case, $\theta=0$, one gets

$$\begin{aligned}\rho_{11} &= \rho_{22} = \frac{1}{2} [\tanh(\beta_1^o + \gamma_o^o) + \tanh(\beta_1^e + \gamma_o^e)] \\ \rho_{12} &= -\rho_{21} = \frac{j}{2} [\tanh(\beta_1^o + \gamma_o^o) - \tanh(\beta_1^e + \gamma_o^e)]. \quad (36)\end{aligned}$$

To relate the transmitted field components at the interface $z=d_m$ to the incident field at $z=0$, we seek expressions of the form

$$\begin{aligned}E_{m+1,o,\perp}^i(d_m) &= (\tau_{ij}^o) E_o^i(0) \\ E_{m+1,e,\perp}^i(d_m) &= (\tau_{ij}^e) E_o^i(0). \quad (37)\end{aligned}$$

At the interface $z=d_m$, we have

$$E_{m+1}^i(d_m) = T^{m,m+1} E_m^i(d_m). \quad (38)$$

At all other interfaces except $z=0$, the fields are related as in (30), and for $z=0$, the relations are given by (32). Upon introduction of appropriate phase factors to relate the fields between any two interfaces, one gets for $\theta=\pi/2$

$$(\tau_{ij}^o) = A^o \begin{pmatrix} \cos^2 \phi & \sin \phi \cos \phi & 0 \\ \sin \phi \cos \phi & \sin^2 \phi & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (39)$$

$$(\tau_{ij}^e) = A^e \begin{pmatrix} \sin^2 \phi & -\sin \phi \cos \phi & 0 \\ -\sin \phi \cos \phi & \cos^2 \phi & 0 \\ j \frac{\eta_{m+1} \sin \phi}{\epsilon_{m+1}} & -j \frac{\eta_{m+1} \cos \phi}{\epsilon_{m+1}} & 0 \end{pmatrix} \quad (40)$$

where

$$\begin{aligned}A^{o,e} &= (1 + \tanh \gamma_o^{o,e}) \exp \left(-j \sum_{k=1}^{k=m} \psi_k^{o,e} \right) \\ &\cdot \prod_{k=1}^{k=m} \left(\frac{1 + \tanh \gamma_k^{o,e}}{1 + \tanh \gamma_{k-1}^{o,e} \tanh \beta_k^{o,e}} \right). \quad (41)\end{aligned}$$

For the case $\theta=0$, one obtains

$$(\tau_{ij}^o) = \frac{A^o}{2} \begin{pmatrix} 1 & j & 0 \\ -j & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (42)$$

$$(\tau_{ij}^e) = \frac{A^e}{2} \begin{pmatrix} 1 & -j & 0 \\ j & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (43)$$

where A^o and A^e have the same form as before.

Sometimes, one is interested in the ratios of power densities. These can be obtained from the above relations by taking time average values of the normal component of the Poynting vector. For the ratios of the reflected and the transmitted power intensities in re-

gions (0) and $(m+1)$, respectively, to the incident power density in region (0) when the incident wave is linearly polarized along the x -axis and the steady magnetic field is in the plane of the interface ($\theta=\pi/2$, ϕ arbitrary), one obtains

$$\begin{aligned}\frac{P_o^r}{P_o^i} &= |\tanh(\beta_1^o + \gamma_o^o)|^2 \cos^2 \phi \\ &+ |\tanh(\beta_1^e + \gamma_o^e)|^2 \sin^2 \phi\end{aligned}$$

$$\frac{P_{m+1}^i}{P_o^i} = \text{Re} \{ n_{m+1}^o |A^o|^2 \cos^2 \phi + n_{m+1}^e |A^e|^2 \sin^2 \phi \}; \quad (44)$$

and when $\theta=0$:

$$\frac{P_o^r}{P_o^i} = \frac{1}{2} \{ |\tanh(\beta_1^o + \gamma_o^o)|^2 + |\tanh(\beta_1^e + \gamma_o^e)|^2 \}$$

$$\frac{P_{m+1}^i}{P_o^i} = \frac{1}{2} \text{Re} \{ n_{m+1}^o |A^o|^2 + n_{m+1}^e |A^e|^2 \}. \quad (45)$$

If the incident wave is circularly polarized, *i.e.*, $E_{oy}^i = \pm j E_{oz}^i$, one gets for $\theta=\pi/2$

$$\frac{P_o^r}{P_o^i} = \frac{1}{2} \{ |\tanh(\beta_1^o + \gamma_o^o)|^2 + |\tanh(\beta_1^e + \gamma_o^e)|^2 \}$$

$$\frac{P_{m+1}^i}{P_o^i} = \frac{1}{2} \text{Re} \{ n_{m+1}^o |A^o|^2 + n_{m+1}^e |A^e|^2 \} \quad (46)$$

and when $\theta=0$:

$$\frac{P_o^r}{P_o^i} = \begin{cases} |\tanh(\beta_1^e + \gamma_o^e)|^2 & \text{when } E_{oy}^i = +j E_{oz}^i \\ |\tanh(\beta_1^o + \gamma_o^o)|^2 & \text{when } E_{oy}^i = -j E_{oz}^i \end{cases} \quad (47)$$

$$\frac{P_{m+1}^i}{P_o^i} = \begin{cases} \text{Re} \{ n_{m+1}^e |A^e|^2 \} & \text{when } E_{oy}^i = +j E_{oz}^i \\ \text{Re} \{ n_{m+1}^o |A^o|^2 \} & \text{when } E_{oy}^i = -j E_{oz}^i \end{cases} \quad (48)$$

In performing the actual calculations of the reflection and transmission coefficients, the procedure is as follows. Starting with the m th interface, one computes β 's and γ 's successively from the recurrence relations of (31) until β_1 is obtained. The procedure is terminated by computing $\tanh \beta_1$ and $\tanh(\beta_1 + \gamma_o)$, where $\gamma_o = \frac{1}{2} \ln(n_1)$ for the ordinary and the extraordinary wave contributions, and putting the result into (35) or (36) to obtain the reflection coefficients; this result is also put in (41) and later into (39) and (40) or (42) and (43) to obtain the transmission coefficients. At each step one is dealing with complex numbers; thus, the procedure may be tedious when a large number of layers is involved. The whole problem, however, may be programmed without difficulty on a modern digital computer.

ACKNOWLEDGMENT

Thanks are extended to Dr. F. S. Holman and T. G. Dalby, both of Boeing Airplane Company, for their advice and encouragement.

The Electromagnetic Field in a Randomly Inhomogeneous Medium

W. C. HOFFMAN†

Summary—Maxwell's equations for a medium in which the "dielectric constant" is a slowly-varying random function of position are reduced to a scalar Helmholtz equation. The Helmholtz equation defines the electromagnetic field components as linear stochastic processes in the random refractive index. Spectral representations for the electromagnetic field components and the refractive index are then introduced, and the former is expressed in terms of a convolution of the latter. The relation between the spectral representations is found explicitly for the case when the average phase is constant.

The case when a source is present is next considered. The inhomogeneous Helmholtz equation is expressed as a singular Fredholm integral equation whose kernel involves the Green's function for the homogeneous medium and the random refractive index. The conditions on the correlation function of the refractive index such that the Neumann series solution converges in the mean-square sense are investigated.

INTRODUCTION

THE theoretical problem of electromagnetic wave motion in a medium which is isotropic but whose index of refraction is a random function of position has been the object of intensive investigation in recent years. A good part of this work has been directed toward the phenomenon of scattering of radio waves by atmospheric turbulence, but the problem possesses an intrinsic importance in its own right and is of interest in several contexts. The theoretical approach to the problem has usually consisted of setting up an integral representation of the electromagnetic field (or a potential associated with it), and after some analysis, time or space averages relating field behavior to refractive index fluctuations have been obtained. That such averages may not be relevant to the problem, and that what are actually required are mathematical expectations or moments, has been pointed out by Kampe de Fariet.¹ An alternative approach based on the fact that the Helmholtz equation can be regarded as defining a linear stochastic process between the random field and refractive index is therefore suggested. A somewhat similar treatment of the simple harmonic oscillator subject to a random forcing function has been carried out by J. M. Burgers,² but there are essential differences between the formulations of the two problems. In Burgers' formulation, the random forcing function does not multiply the unknown function. In the random medium case, it is necessary to transform the Helmholtz equation to a multidimensional analog of the Riccati equation

in order to obtain such a form. As is the case with the ordinary Riccati equation, the resulting differential equation is nonlinear, and the transformation appears to be of value only in certain special cases, as will be set forth in more detail below.

Maxwell's equations lead to scalar Helmholtz equations in the field components only if the change in refractive index of the medium over the distance of one wavelength is small. Otherwise, the electric field components are coupled to each other and to the components of the magnetic field vector. The general coupled equations are first derived. The hypothesis $|\nabla n^2|\lambda \ll 1$ is next invoked and the resulting Helmholtz equation is treated in two ways. First, a transformation to a multidimensional analog of the Riccati equation is carried out, a perturbation analysis is then applied to this differential equation, and spectral representations for the first perturbations of field and refractive index are introduced. This permits the determination of the relation between the associated orthogonal random processes.

The multidimensional Riccati equation leads to useful results only if the zero-order perturbation of the phase has a particularly simple form. In other cases it appears better to use the original Helmholtz equation and introduce the spectral representations of field vector and refractive index at the outset. An integral equation between the respective stochastic processes is thereby obtained.

Finally, the case when a source is present but $|\nabla n^2|\lambda$ is still $\ll 1$, so that we have to deal with an inhomogeneous Helmholtz equation, is taken up. The inhomogeneous Helmholtz equation is reduced to a singular Fredholm integral equation, and a solution in the form of a Neumann series is obtained. The latter is a generalized form of the usual Neumann series (or resolvent kernel) solution, and the conditions for its existence and convergence are determined.

THE CASE OF A SOURCE-FREE RANDOMLY INHOMOGENEOUS MEDIUM

The General Coupled Equations

It will first be supposed that the index of refraction varies only very slowly as a function of time, so that its time derivative effectively vanishes over the period under consideration.³ Simple harmonic time dependence

† University of Queensland, Brisbane, Australia. Formerly at Hughes Research Laboratories, Culver City, Calif.

¹ J. Kampe de Fariet, "Spectral Tensor of Homogeneous Turbulence," presented at the Symposium on Turbulence, Naval Ord. Lab. Rept. No. 1136; July 1, 1949.

² G. K. Batchelor, "The Theory of Homogeneous Turbulence," Cambridge University Press, Cambridge, Eng., sect. 4.1; 1953.

³ Alternatively, one could restrict attention to independent realizations. A nonnegligible time derivative has an effect similar to that of a nonvanishing conductivity of the medium, for Ampere's equation becomes $\nabla \wedge \mathbf{H} = i\omega \mathbf{E} + \dot{\epsilon} \mathbf{E}$, the right-hand side of which is formally similar to $(i\omega\epsilon + \sigma) \mathbf{E}$ for nonzero $\partial\epsilon/\partial t$.

$e^{i\omega t}$ may then be assumed, and Maxwell's equations for a dielectric medium take on the form

$$\nabla \wedge \mathbf{E} = -i\omega\mu_0\mathbf{H} \quad \nabla \cdot \mathbf{E} = -\nabla[\ln n^2(\mathbf{x}) \cdot \mathbf{E}] \quad (1)$$

$$\nabla \wedge \mathbf{H} = i\omega\epsilon_0 n^2(\mathbf{x})\mathbf{E} \quad \nabla \cdot \mathbf{H} = 0, \quad (2)$$

where μ_0 , ϵ_0 are the constitutive parameters of free space, $n^2(\mathbf{x})$ is the refractive index, and \mathbf{x} denotes a point in 3-space with components (x_1, x_2, x_3) . Eqs. (1) and (2) lead in the usual way to the coupled system of vector equations:

$$\begin{aligned} \nabla^2 \mathbf{E} + k_0^2 [\mathfrak{M}n^2(\mathbf{x}) + \nu^2(\mathbf{x})] \mathbf{E} \\ = -\nabla(\nabla \ln [\mathfrak{M}n^2(\mathbf{x}) + \nu^2(\mathbf{x})] \cdot \mathbf{E}) \\ \nabla^2 \mathbf{H} + k_0^2 [\mathfrak{M}n^2(\mathbf{x}) + \nu^2(\mathbf{x})] \mathbf{H} \\ = -i\omega\epsilon_0 [\nabla[\mathfrak{M}n^2(\mathbf{x}) + \nu^2(\mathbf{x})]] \wedge \mathbf{E}, \quad (3) \end{aligned}$$

where $\mathfrak{M}n^2(\mathbf{x})$ denotes the mathematical expectation of $n^2(\mathbf{x})$ and $\nu^2(\mathbf{x})$ is the random residual

$$\nu^2(\mathbf{x}) = n^2(\mathbf{x}) - \mathfrak{M}n^2(\mathbf{x}). \quad (4)$$

It is clear that not only are the equations coupled to one another by the right-hand sides, but that the various components of the field vectors are coupled to each other. If we now regard $n^2(\mathbf{x})$ as a stationary random function of position and note that (3) then defines the electromagnetic field vectors \mathbf{E} and \mathbf{H} as random functions in turn, we may introduce such spectral representations as

$$\mathbf{E}(\mathbf{x}) = \int_{R_3} e^{i\alpha \cdot \mathbf{x}} d\mathcal{E}(\alpha), \quad (5)$$

etc., where $\mathcal{E}(\alpha)$ is a stochastic process whose exact nature will be defined later, and the integration is extended over three-dimensional Euclidean space R_3 . One again finds two coupled vector equations in terms of the spectral representations, which offer no particular advantage in tractability over the original system (3).

We therefore consider the case when the refractive index varies not only slowly in time but also very gradually in space. More precisely *the gradient of refractive index over the distance of one wavelength is small:*

$$|\nabla n^2| \lambda \ll 1. \quad (6)$$

Under this hypothesis, the right-hand sides of (3) may be neglected.^{4,5} This uncouples both wave equations and field vectors, and permits us to write any one of the six scalar Helmholtz equations which result from (3) in terms of any particular component, thus

$$\nabla^2 F + k_0^2 [\mathfrak{M}n^2(\mathbf{x}) + \nu^2(\mathbf{x})] F(\mathbf{x}) = 0, \quad (7)$$

⁴ J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., Inc., New York, N. Y., p. 343; 1941.

⁵ K. A. Suchy, "Lösung der Gleichungen für Wellennormale und Brechungsindex durch WBK-Näherung. Strahlenoptische Reflexion und Alternation," *Ann. der Physik*, vol. 13, pp. 178-197, sect. 28, pp. 194-196; 1953. Suchy's form of the condition reads

$$|\nabla \nabla \cdot \mathbf{E}|_{\mathbf{e}_k} \ll k_0^2 \left| \frac{\tilde{\epsilon}}{\epsilon_0} \mathbf{E} \right|.$$

where $F(\mathbf{x})$ denotes some component of $\mathbf{E}(\mathbf{x})$ or $\mathbf{H}(\mathbf{x})$. We may now proceed in one of two ways, either directly to the spectral representations of the field quantities or by first transforming the Helmholtz equation (7) to a multidimensional Riccati equation. Since the latter course leads to more immediate results, we follow it first.

Transformation of the Helmholtz Equation. Perturbation Solution

According to the discussion of the previous section, the system (3) can be replaced by two or more scalar Helmholtz equations of the form (7). The transformation

$$F(\mathbf{x}) = a(\mathbf{x}) \exp \{ -ik_0 S(\mathbf{x}) \},$$

or

$$F(\mathbf{x}) = a_0 \exp \{ -ik_0 \psi(\mathbf{x}) \}, \quad (8)$$

with the constant amplitude a_0 and the generalized phase function

$$\psi(\mathbf{x}) = S(\mathbf{x}) + \frac{i}{k_0} \ln (a(\mathbf{x})/a_0), \quad (9)$$

transforms (7) into the nonlinear equation

$$(\nabla \psi)^2 + \frac{i}{k_0} \nabla^2 \psi = \mathfrak{M}n^2(\mathbf{x}) + \nu^2(\mathbf{x}). \quad (10)$$

If we now define

$$\nabla \psi = \mathbf{X}, \quad (11)$$

we see that (10) is the spatial analog of the Riccati equation⁶ associated with the one-dimensional Helmholtz equation, for (10) then reads

$$\nabla \cdot \mathbf{X} - ik_0 \mathbf{X}^2 = -ik_0 [\mathfrak{M}n^2(\mathbf{x}) + \nu^2(\mathbf{x})]. \quad (12)$$

We now suppose that the refractive index $n^2(\mathbf{x})$ has an expansion in terms of a parameter $\delta \ll 1$:

$$n^2(\mathbf{x}) = \mathfrak{M}n^2(\mathbf{x}) + \sum_{j=1}^{\infty} \nu_j(\mathbf{x}) \delta^j. \quad (13)$$

It follows from the form (13) that to terms of higher order than δ , $n^2(\mathbf{x})$ can be regarded as a random function with mean $\mathfrak{M}n^2(\mathbf{x})$ and standard deviation δ , i.e.,

⁶ The one-dimensional Helmholtz equation, $u''(x) + p(x)u(x) = 0$, goes into the Riccati equation, $y'(x) + y^2 = -p(x)$, under the transformation $u(x) = \exp \{ \int y dx \}$. If $p(x)$ is an analytic random function (see Moyal, footnote 7) whose expansion is

$$p(x) = \sum_0^{\infty} p_j x^j, \quad \text{then} \quad y(x) = \sum_0^{\infty} c_j x^j,$$

and the following recurrence relation for the expansion coefficients c_j , ($j=0, 1, 2, \dots$), may be obtained from the Riccati equation,

$$j c_j = - \sum_{h=0}^{j-1} c_h c_{j-1-h} - p_j.$$

Knowledge of the (nonrandom) coefficient c_0 is therefore enough to specify the wave function in this case.

⁷ J. E. Moyal, "Stochastic processes and statistical physics," *J. Roy. Stat. Soc. (B)*, vol. 11, pp. 150-210; 1949.

$\nu_1(\mathbf{x})$ is a standardized random variable with mean zero and unit variance. It is next assumed that the gradient $\kappa(\mathbf{x})$ of the generalized phase function $\psi(\mathbf{x})$ also possesses an asymptotic expansion, convergent in a neighborhood of $\delta=0$, of the form

$$\chi(\mathbf{x}) \sim \sum_{j=0}^{\infty} \chi_j(\mathbf{x}) \delta^j. \quad (\delta \ll 1). \quad (14)$$

Substitution of (13) and (14) into the differential equation (10) and equating coefficients of j th powers of δ on each side of the resulting equation yields the system,

$$\begin{aligned} \frac{i}{k_0} \nabla \cdot \kappa_0 + \kappa_0^2 &= \mathfrak{M}n^2 \\ \frac{i}{k_0} \nabla \cdot \kappa_j + \sum_{h=0}^j \kappa_h \cdot \kappa_{j-h} &= \nu_j(\mathbf{x}), \quad (j = 1, 2, \dots) \end{aligned} \quad (15)$$

in which the equations are now linear for $j > 0$. This seems to be the closest analog of Burgers' approach.² The zero-order equation has already been written out explicitly in the first equation of (15); the first-order perturbation equation is

$$\frac{i}{k_0} \nabla \cdot \kappa_1 + 2\kappa_0 \cdot \kappa_1 = \nu_1(\mathbf{x}). \quad (16)$$

The right-hand side of the zero-order equation may be a function of position or it may be constant, but in any case, $\mathfrak{M}n^2(\mathbf{x})$ is a sure (*i.e.*, a nonrandom) function. The right-hand side of (16) is of course a random function. In the event that $\mathfrak{M}n^2(\mathbf{x})$ is constant or of the form $\mathfrak{M}n^2(\mathbf{x}) = f_1(x_1) + f_2(x_2) + f_3(x_3)$, the zero-order equation can be solved by separation of variables in the same way as the classical eikonal equation. An important feature of the asymptotic expansion (14) is the separation of $\kappa(\mathbf{x})$ into sure and random components, the sure component being the solution for a possibly inhomogeneous, but certainly nonrandom, medium.

If $\psi(\mathbf{x})$ does have an asymptotic expansion of the form (14) then, to terms of higher order than δ^2 , the covariance function of $\kappa(\mathbf{x})$ is proportional to that of $\kappa_1(\mathbf{x})$; *i.e.*,

$$\text{cov} [\kappa(\mathbf{x}), \kappa(\mathbf{x}')] = \delta^2 \text{cov} [\kappa_1(\mathbf{x}), \kappa_1(\mathbf{x}')] + o(\delta^2). \quad (17)$$

This result follows immediately upon computing

$$\begin{aligned} \text{cov} [\kappa(\mathbf{x}), \kappa(\mathbf{x}')] &= \mathfrak{M} \{ \kappa(\mathbf{x}) \cdot \kappa^*(\mathbf{x}') \} - \mathfrak{M} \{ \kappa(\mathbf{x}) \} \cdot \mathfrak{M} \{ \kappa^*(\mathbf{x}') \}, \end{aligned}$$

introducing the expansion (14) and making use of the fact (from the zero-order equation) that $\kappa_0(\mathbf{x})$ is a sure function if $\mathfrak{M}n^2(\mathbf{x})$ is.

The Case of Constant Gradient of ψ_0 .

Whenever

$$\nabla \psi_0 = \kappa_0 = \text{const.}, \quad (18)$$

the zero-order equation is trivially satisfied. Physically speaking, this corresponds to the situation when the medium has no variation of refractive index other than

turbulent fluctuations. In general, the refractive index of an actual atmosphere will consist of turbulent fluctuations superimposed on a slower systematic variation due to gravity, large-scale meteorological phenomena, etc. Consideration of this case is postponed to the next section.

It is physically reasonable that the first perturbation $\nu_1(\mathbf{x})$ of the refractive index is a stationary (or spatially homogeneous) stochastic process with mean zero. Under this hypothesis the first perturbation has a spectral representation,⁸

$$\nu_1(\mathbf{x}) = \int_{R_3} e^{i\alpha \cdot \mathbf{x}} d\eta_1(\alpha), \quad (19)$$

where R_3 denotes the three-dimensional Euclidean space of points $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ and $\eta_1(\alpha)$ is a stochastic process with orthogonal increments. Since κ_0 is constant according to the hypothesis, $\kappa_1(\mathbf{x})$ is also a stationary random process, with a spectral representation in terms of an orthogonal process. We shall assume that this spectral representation is of the form

$$\kappa_1(\mathbf{x}) = \int_{R_3} e^{i\alpha \cdot \mathbf{x}} d\xi_1(\alpha) \gamma_1(\alpha), \quad (20)$$

where $\xi_1(\alpha)$ is a process with orthogonal increments and $\gamma_1(\alpha)$ is a sure vector function of α .

The Fourier transform of (16) then yields, in view of the uniqueness of the transform,⁸ and the hypothesis (18) and definitions (19) and (20):

$$d\xi_1(\alpha) = k_0 \frac{d\eta_1(\alpha)}{2k_0\kappa_0 \cdot \gamma_1(\alpha) - \alpha \cdot \gamma_1(\alpha)}. \quad (21)$$

Relation (21) connects the spectral distribution functions of the first perturbations of the field and gradient of refractive index.

If we now form

$$d\Phi_1(\alpha) = \mathfrak{M} \{ d\xi_1(\alpha) d\xi_1^*(\alpha') \}, \quad (22)$$

and let

$$d\sigma_1(\alpha) = \mathfrak{M} \{ d\eta_1(\alpha) d\eta_1^*(\alpha') \}, \quad (23)$$

there then exists, in the usual way, the following relation between the energy spectrum of the κ_1 -process and that of the first-order refractive index fluctuations:

$$d\Phi_1(\alpha) = k_0^2 \frac{d\sigma_1(\alpha)}{[2k_0\kappa_0 \cdot \gamma_1(\alpha) - \alpha \cdot \gamma_1(\alpha)]^2}. \quad (24)$$

The covariance function of κ_1 can then be found as the Fourier transform of $d\Phi_1(\alpha)$:

$$\begin{aligned} R_1(\mathbf{x} - \mathbf{x}') &= \mathfrak{M} \{ \kappa_1(\mathbf{x}) \cdot \kappa_1^*(\mathbf{x}') \} \\ &= k_0^2 \int_{R_3} e^{i(\mathbf{x} - \mathbf{x}') \cdot \alpha} \frac{d\sigma_1(\alpha)}{[2k_0\kappa_0 \cdot \gamma_1(\alpha) - \alpha \cdot \gamma_1(\alpha)]^2}. \end{aligned} \quad (25)$$

⁸ A. Blanc-Lapierre and R. Fortet, "Theorie des Fonctions Aleatoires," Masson, Paris, France, pp. 363-364; 1953.

Since the covariance function $R_1(\mathbf{x} - \mathbf{x}')$ of χ_1 is the Fourier transform of a product, it is representable as the convolution of the inverse transforms of $d\sigma_1(\alpha)$ and

$$[2k_0\chi_0 \cdot \gamma_1(\alpha) - \alpha \cdot \gamma_1(\alpha)]^{-2}. \quad (26)$$

Thus

$$R_1(\mathbf{x} - \mathbf{x}') = \int_{R_3} \rho_1(\mathbf{x} - \mathbf{x}' - \mathbf{w}) d g_1(\mathbf{w}), \quad (27)$$

where

$$\rho_1(\mathbf{x} - \mathbf{x}') = \int_{R_3} e^{i(\mathbf{x} - \mathbf{x}') \cdot \alpha} d\sigma_1(\alpha), \quad (28)$$

and $g_1(\mathbf{w})$ is the inverse Fourier transform of (26), if it exists.

Case of Nonconstant $\nabla\psi_0$

When $\nabla\psi_0 = \chi_0$ is not constant but a function of position (which is the more realistic situation since actual atmospheres ordinarily have a refractive index made up of a gradual over-all large-scale variation with small local fluctuations superimposed upon this), (15) no longer appears to be useful. It seems to be more convenient to return to the original Helmholtz equation (7) and introduce the spectral representations of refractive index and field component at that stage. The refractive index may still be regarded as a stationary process (at least in the wide sense), but it is not immediately clear that the field can, and a new basis must be found for its spectral representation. This is available in a theorem⁸ which asserts the existence of the spectral representation

$$F(\mathbf{x}) = \int_{R_3} e^{i\mathbf{x} \cdot \alpha} d\zeta(\alpha), \quad (29)$$

$\zeta(\alpha)$ being a stochastic process such that

$$\text{cov} [F(\mathbf{x}), F(\mathbf{x}')] = \int_{R_3 \times R_3} e^{i(\mathbf{x} \cdot \alpha - \mathbf{x}' \cdot \beta)} d^2\Gamma(\alpha, \beta) \quad (30)$$

with

$$d^2\Gamma(\alpha, \beta) = \Re \{ d\zeta(\alpha) d\zeta^*(\beta) \}, \quad (31)$$

and $d^2\Gamma(\alpha, \beta)$ is absolutely integrable over the whole space R_3 . Thus the covariance function must be uniformly bounded and continuous over the product space $R_3 \times R_3$, which in turn implies the mean square continuity of $F(\mathbf{x})$. The random part of the refractive index, being a second-order stationary process, has a spectral representation

$$\nu^2(\mathbf{x}) = \int_{R_3} e^{i\mathbf{x} \cdot \alpha} d\eta(\alpha), \quad (32)$$

where $\eta(\alpha)$ is a process with orthogonal increments. We further suppose that the average refractive index

$\Re n^2(\mathbf{x})$ can be represented as a Fourier integral:

$$\Re n^2(\mathbf{x}) = \int_{R_3} e^{i\mathbf{x} \cdot \alpha} m(\alpha) d\alpha. \quad (33)$$

The Fourier transform of (7) then yields the following integral equation connecting the various spectral functions:

$$(\alpha \cdot \alpha) d\zeta(\alpha) = k_0^2 \int_{R_3} \{ m(\alpha - \beta) + d\eta(\alpha - \beta) \} d\zeta(\beta). \quad (34)$$

This integral equation is formally similar to the integral equation of the third kind,⁹ but the theory of such integral equations would require considerable extension in order to handle a random integral equation such as (34). The usual way of solving an integral equation such as (34) is to invoke the Fourier integral theorem. However, such a course would, in the present instance, simply lead us back along the route we have come.

THE INHOMOGENEOUS HELMHOLTZ EQUATION FOR A RANDOMLY INHOMOGENEOUS MEDIUM

Reduction to a Fredholm Equation

When a source of current \mathbf{J} and charge density ρ are also present, the Maxwell equations (1) and (2) lead, under the condition expressed by (6), to the following inhomogeneous Helmholtz equation in the electromagnetic field components:

$$\nabla^2 F + k_0 n^2(\mathbf{x}) F = f(\mathbf{x}), \quad (35)$$

defined over the (possibly infinite) domain D .

Introduce the new function

$$q(\mathbf{x}) = 1 - n^2(\mathbf{x}), \quad (36)$$

and the Green's function $G(\mathbf{x}, \mathbf{x}')$ for the homogeneous medium:

$$\nabla_{\mathbf{x}}^2 G(\mathbf{x}, \mathbf{x}') + k_0^2 G(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}'), \quad (37)$$

with homogeneous boundary conditions for G or its normal derivative. Then, in the usual way, we obtain the following integral equation for $F(\mathbf{x})$:

$$\begin{aligned} \int_D \{ F(\mathbf{x}') \nabla_{\mathbf{x}'}^2 G(\mathbf{x}, \mathbf{x}') - G(\mathbf{x}, \mathbf{x}') \nabla_{\mathbf{x}'}^2 F(\mathbf{x}') \} d\mathbf{x}' \\ = \int_D \{ \delta(\mathbf{x} - \mathbf{x}') F(\mathbf{x}') - k_0^2 G(\mathbf{x}, \mathbf{x}') q(\mathbf{x}') F(\mathbf{x}') \\ - G(\mathbf{x}, \mathbf{x}') f(\mathbf{x}') \} d\mathbf{x}'. \end{aligned} \quad (38)$$

The integrals on the left-hand side of (38) will be assumed to exist in the mean-square sense. (Doob¹⁰ gives conditions for the existence of such stochastic integrals

⁹ W. Schmeidler, "Integralgleichungen mit Anwendungen in Physik und Technik I," 2nd ed., Akad. Verlag, Leipzig, Germany, ch. 4; 1955.

¹⁰ J. L. Doob, "Stochastic Processes," John Wiley and Sons, New York, N. Y., p. 430; 1953.

in the case that the random part of the integrand constitutes a process with orthogonal increments.) Application of Green's theorem (in the mean-square sense) then yields

$$\int_{\text{bdry } D} \left\{ F(\mathbf{x}') \frac{\partial G}{\partial n} - G(\mathbf{x}, \mathbf{x}') \frac{\partial F}{\partial n} \right\} dS_{\mathbf{x}'} = F(\mathbf{x}) - k_0^2 \int_P G(\mathbf{x}, \mathbf{x}') q(\mathbf{x}') F(\mathbf{x}') d\mathbf{x}' - g(\mathbf{x}), \quad (39)$$

where

$$g(\mathbf{x}) = \int_D G(\mathbf{x}, \mathbf{x}') f(\mathbf{x}') d\mathbf{x}'. \quad (40)$$

Homogeneous Dirichlet or Neumann boundary conditions will be assumed as well as a radiation condition if some or all of boundary D is at infinity. The left-hand side of (39) then vanishes, and we obtain finally the inhomogeneous Fredholm integral equation

$$F(\mathbf{x}) = g(\mathbf{x}) + k_0^2 \int_D G(\mathbf{x}, \mathbf{x}') q(\mathbf{x}') F(\mathbf{x}') d\mathbf{x}'. \quad (41)$$

The integral (41) may be singular either in that the domain D is infinite, or that the kernel is singular. However, we assume that in any case the integral on the right exists in the sense of mean-square convergence, i.e., while G may exhibit a logarithmic singularity or D may be

Form the (formal) Neumann series:

$$\begin{aligned} N(\mathbf{x}) = & g(\mathbf{x}) + k_0^2 \int_D G(\mathbf{x}, \mathbf{x}') q(\mathbf{x}') g(\mathbf{x}') d\mathbf{x}' + \dots \\ & + k_0^{2m} \int_D d\mathbf{x}_1 G(\mathbf{x}, \mathbf{x}_1) q(\mathbf{x}_1) \int_D d\mathbf{x}_2 \dots \\ & \cdot \int_D d\mathbf{x}_{m-1} G(\mathbf{x}_{m-2}, \mathbf{x}_{m-1}) q(\mathbf{x}_{m-1}) \\ & \cdot \int_D G(\mathbf{x}_{m-1}, \mathbf{x}_m) q(\mathbf{x}_m) g(\mathbf{x}_m) d\mathbf{x}_m + \dots \end{aligned} \quad (42)$$

We now add the hypothesis that $n^2(\mathbf{x})$ is a Gaussian process with mean 1 and variance v . It then follows from (36) that $q(\mathbf{x})$ is Gaussian with mean zero and variance v . We may then use the formula (which can be readily derived from the characteristic function for a multivariate normal distribution)

$$\mathfrak{M}\{x_1 x_2 \dots x_n\} = \sum_{s_j \neq s_k} r_{s_1 s_2} r_{s_3 s_4} \dots r_{s_{n-1} s_n}, \quad (n \text{ even}) \quad (43)$$

(2) terms.

where $r_{s_j s_k}$ denotes the covariance of x_{s_j} and x_{s_k} , in order to establish the mean-square convergence of the right hand side of (42).

The $(m+1)$ st term on the right in (42) has the mean-square value

$$\begin{aligned} \mathfrak{M} \left\{ \left| k_0^{2m} \int_D d\mathbf{x}_1 G(\mathbf{x}, \mathbf{x}_1) q(\mathbf{x}_1) \int_D d\mathbf{x}_2 \dots \int_D G(\mathbf{x}_{m-1}, \mathbf{x}_m) q(\mathbf{x}_m) g(\mathbf{x}_m) d\mathbf{x}_m \right|^2 \right\} \\ = \left| k_0^2 \right|^{2m} \int_D d\mathbf{x}_1 G(\mathbf{x}, \mathbf{x}_1) \int_D d\mathbf{x}_2 \dots \int_D d\mathbf{x}_m G(\mathbf{x}_{m-1}, \mathbf{x}_m) q(\mathbf{x}_m) g(\mathbf{x}_m) \int_D d\mathbf{x}_1' G^*(\mathbf{x}, \mathbf{x}_1') \int_D d\mathbf{x}_2' \dots \\ \cdot \int_D d\mathbf{x}_m' G^*(\mathbf{x}_{m-1}', \mathbf{x}_m') \mathfrak{M}\{q(\mathbf{x}_1) \dots q(\mathbf{x}_m) q^*(\mathbf{x}_1') \dots q^*(\mathbf{x}_m')\}. \end{aligned} \quad (44)$$

infinite in extent, $\mathfrak{M}\{|q(\mathbf{x}') F(\mathbf{x}')|^2\}$ behaves in such a way that the integral itself is finite.

The Neumann Series Solution

As just mentioned, the kernel G or the domain D may be such that the integral equation is singular. In that case we cannot in general interchange the order of integration as required in establishing the usual Neumann series solution. However, we can still form a formal Neumann series solution and show its convergence to the solution of (41). This is the procedure usually followed in the nonrandom singular case,¹¹ and we shall show that it suffices in the random case also, provided the Neumann series solution for the homogeneous medium with source converges in the usual sense.

By (43), the last factor of the integrand of (44) can be written as the sum of $\binom{2m}{2}$ products of all possible covariances of the $q(\mathbf{x}_j)$ and $q^*(\mathbf{x}_j')$, ($j=1, \dots, m$). Since $r_{s_j s_j} \leq \text{var } q^2 = v$, we thus have the following bound for (44):

$$\begin{aligned} \mathfrak{M} \left\{ \left| k_0^{2m} \int_D d\mathbf{x}_1 G(\mathbf{x}, \mathbf{x}_1) q(\mathbf{x}_1) \int_D d\mathbf{x}_2 \dots \right. \right. \\ \left. \cdot \int_D G(\mathbf{x}_{m-1}, \mathbf{x}_m) q(\mathbf{x}_m) g(\mathbf{x}_m) d\mathbf{x}_m \right|^2 \} \\ \leq \binom{2m}{2} \left| k_0^4 v \right|^m \left| \int_D d\mathbf{x}_1 G(\mathbf{x}, \mathbf{x}_1) \int_D d\mathbf{x}_2 \dots \right. \\ \left. \cdot \int_D G(\mathbf{x}_{m-1}, \mathbf{x}_m) g(\mathbf{x}_m) d\mathbf{x}_m \right|^2. \end{aligned} \quad (45)$$

¹¹ W. Schmeidler, *ibid.*, sect. 23(a).

Now the last factor on the right in (45) corresponds to the bound on the $(m+1)$ st term in the ordinary Neumann series development, so that the problem of convergence is essentially reduced to that of the non-random case. We therefore assume that the Neumann series for the ordinary homogeneous medium case converges in the usual way, either because the integral equation

$$U(\mathbf{x}) = g(\mathbf{x}) + k_0^2 \int_D G(\mathbf{x}, \mathbf{x}') U(\mathbf{x}') d\mathbf{x}'$$

is nonsingular, or because $\int_D G(\mathbf{x}, \mathbf{x}') g(\mathbf{x}') d\mathbf{x}'$ and its iterates are such that the integral on the right in (45) converges. In that case, we may write

$$\begin{aligned} \mathfrak{N} \left\{ \left| k_0^{2m} \int_D d\mathbf{x}_1 G(\mathbf{x}, \mathbf{x}_1) q(\mathbf{x}_1) \int_D d\mathbf{x}_2 \cdots \right. \right. \\ \left. \left. \cdot \int_D G(\mathbf{x}_{m-1}, \mathbf{x}_m) q(\mathbf{x}_m) g(\mathbf{x}_m) d\mathbf{x}_m \right|^2 \right\} \\ \leq \binom{2m}{2} |k_0^4 v M^2|^m. \end{aligned} \quad (46)$$

It then follows from the Minkowski inequality applied to the square of (42) that

$$\mathfrak{N}^{1/2} \{ |N(\mathbf{x})|^2 \} < |g(\mathbf{x})| + \sqrt{2} \frac{|k_0^2 v^{1/2} M|}{(1 - |k_0^2 v^{1/2} M|)^2},$$

provided

$$|k_0^2 v^{1/2} M| < 1. \quad (47)$$

This establishes the convergence of the Neumann series (42).

We have yet to demonstrate that the Neumann series (42) converges in mean square to a solution of (41). To this end, form the partial sum

$$\begin{aligned} N_P(\mathbf{x}) = g(\mathbf{x}) + \sum_{m=1}^P k_0^{2m} \int_D d\mathbf{x}_1 G(\mathbf{x}, \mathbf{x}_1) q(\mathbf{x}_1) \int_D d\mathbf{x}_2 \cdots \\ \cdot \int_D G(\mathbf{x}_{m-1}, \mathbf{x}_m) q(\mathbf{x}_m) g(\mathbf{x}_m) d\mathbf{x}_m, \end{aligned} \quad (48)$$

and apply the Cauchy convergence criterion (which is equivalent to demonstrating that $\lim_{P \rightarrow \infty} \mathfrak{N} \{ |F(\mathbf{x})$

$$- N_P(\mathbf{x})|^2 \} \rightarrow 0):$$

$$\begin{aligned} \mathfrak{N} \{ |N_{P+Q}(\mathbf{x}) - N_P(\mathbf{x})|^2 \} \\ = \mathfrak{N} \left\{ \left| \sum_{m=P+1}^{P+Q} k_0^{2m} \int_D d\mathbf{x}_1 G(\mathbf{x}, \mathbf{x}_1) q(\mathbf{x}_1) \int_D d\mathbf{x}_2 \cdots \right. \right. \\ \left. \left. \cdot \int_D G(\mathbf{x}_{m-1}, \mathbf{x}_m) q(\mathbf{x}_m) g(\mathbf{x}_m) d\mathbf{x}_m \right|^2 \right\}. \end{aligned} \quad (49)$$

By the Minkowski inequality the right-hand side is less than the square of

$$\begin{aligned} \sqrt{2} \frac{|k_0^2 v^{1/2} M|^P}{|k_0^2 v^{1/2} M| - 1} \left\{ (Q+1) |k_0^2 v^{1/2} M|^Q \right. \\ \left. + P |k_0^2 v^{1/2} M| (|k_0^2 v^{1/2} M|^Q - 1) \right. \\ \left. - \frac{|k_0^2 v^{1/2} M|^{Q+1} - 1}{|k_0^2 v^{1/2} M| - 1} \right\}, \end{aligned} \quad (50)$$

and according to (47), the bound (50) converges to zero as $P \rightarrow \infty$. Hence the mean square convergence of $N(\mathbf{x})$ to $F(\mathbf{x})$ is established, in view of the previous definition of $N(\mathbf{x})$ as the (formal) iterative solution of (41). We may sum up in the following theorem:

Theorem. The boundary value problem for

$$\nabla^2 F(\mathbf{x}) + k_0^2 n^2(\mathbf{x}) F(\mathbf{x}) = f(\mathbf{x})$$

with homogeneous Dirichlet or Neumann boundary conditions and a radiation condition if some part of the boundary of the domain D is at infinity, is equivalent to the inhomogeneous Fredholm integral equation

$$F(\mathbf{x}) = g(\mathbf{x}) + k_0^2 \int_D G(\mathbf{x}, \mathbf{x}') q(\mathbf{x}') F(\mathbf{x}') d\mathbf{x}', \quad (51)$$

where $q(\mathbf{x}) = 1 - n^2(\mathbf{x})$, provided the integral and the equality in (51) exist in the mean-square sense.

If

(i) $n^2(\mathbf{x})$ is a Gaussian process $(1, v)$;

and

(ii) $|k_0^2 v^{1/2} M| < 1$, where M is a bound for

$$\left| \int_D G(\mathbf{x}, \mathbf{x}') g(\mathbf{x}') d\mathbf{x}' \right|^2,$$

and

(iii) *the Neumann series solution for the integral equation*

$$U(\mathbf{x}) = g(\mathbf{x}) + k_0^2 \int_D G(\mathbf{x}, \mathbf{x}') U(\mathbf{x}') d\mathbf{x}'$$

converges in the ordinary way,

then the integral equation (51) is solved by the mean-square convergent Neumann series

$$\begin{aligned} F(\mathbf{x}) = g(\mathbf{x}) + \sum_{m=1}^{\infty} k_0^{2m} \int_D d\mathbf{x}_1 G(\mathbf{x}, \mathbf{x}_1) q(\mathbf{x}_1) \int_D d\mathbf{x}_2 \cdots \\ \cdot \int_D G(\mathbf{x}_{m-1}, \mathbf{x}_m) q(\mathbf{x}_m) g(\mathbf{x}_m) d\mathbf{x}_m. \end{aligned}$$

ACKNOWLEDGMENT

Valuable discussions with R. F. Goodrich, M. L. Juncosa, and A. J. F. Siegert are gratefully acknowledged.

Scattering by Quasi-Periodic and Quasi-Random Distributions*

V. TWERSKY†

Summary—We consider the scattering of plane electromagnetic waves by parallel, coplanar, arbitrary cylinders distributed essentially as in a “one-dimensional liquid” of elastic objects. Green’s function methods are used to generalize and extend results obtained previously¹ by separations of variables for circular cylinders. Taking into account coherent multiple scattering, we obtain a general form for the coherent field and a corresponding approximation for the incoherent scattering. The field depends critically on the normalized difference between the average and minimum separations of scatterer centers; say on $d = (b_{av} - b_{min})/b_{av}$ which equals the relative “elbow room” per scatterer. If $d = 0$, then the distribution is periodic and the results reduce to those for the general grating;² thus the range $d \approx 0$ corresponds to the quasi-periodic case. Similarly, at the other limit $d \rightarrow 1$, the results reduce to those for the random “rare gas” case,³ and the range $d \approx 1$ may be called quasi-random. Thus, as the parameter d is varied from 1 to 0 (or as the distribution of scatterers is “compressed”), the result exhibit successively the effects expected for gaseous, liquid, and crystalline distributions.

I. INTRODUCTION

MULTIPLE scattering treatments for densely packed, quasi-random, and quasi-periodic distributions of arbitrary objects are required to close the gap between the relatively well-known gas-like and periodic limits. The present results generalize and extend those obtained previously for a “one-dimensional liquid” of circular cylinders,¹ and thereby provide a continuous transition formalism between the one-dimensional “crystal”² and “gas”³ of arbitrary cylindrical scatterers. The liquid is specified by a Poisson one-particle distribution function, and by a more convergent transform of the pair distribution function introduced by Zernike and Prins.⁴

The present paper is based on the multiple scattering formalism of references 2 and 3, and essentially applies the averaging procedure followed in reference 3 to the case of a more general distribution; the earlier paper³ is to be consulted for the introductory discussion and mathematical formulation of the problem.

* This work was supported in part by Signal Corps Contract DA 36-309 SC 78281. Additional details are given in Sylvania Electronic Defense Lab. Rept. EDL-E37; 1959.

† Sylvania Electronic Defense Lab., Mountain View, Calif.
¹ V. Twersky, “Multiple Scattering of Waves by Planar Random Distribution of Cylinders and Bosses,” Inst. of Mathematical Sciences, New York University, New York, N. Y., Rept. EM 58; October, 1953. Also “A Preliminary Report on the Reflection of Waves from Planar Distributions of Parallel Circular Mirrors,” Nuclear Dev. Assoc., New York, N. Y., Rept. NDA 18-3; August, 1952.

² V. Twersky, “On the scattering of waves by an infinite grating,” IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-4, pp. 330-345; July, 1956.

³ V. Twersky, “On scattering and reflection of sound by rough surfaces,” *J. Acoust. Soc. Amer.*, vol. 29, pp. 209-225; February, 1957. Also, “On scattering and reflection of the electromagnetic waves by rough surfaces,” IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-5, pp. 81-90; January, 1957. Reference 3 as used in the text applies only to the first of these papers. The initial formalism is essentially an extension of that introduced by Foldy to treat a “rare gas” of point scatterers. L. L. Foldy, *Phys. Rev.*, vol. 67, p. 107; 1945.

⁴ F. Zernike and J. A. Prins, *Z. Physik*, vol. 41, p. 184; 1927.

We write the field of an arbitrary planar configuration of N parallel arbitrary cylinders excited by a plane wave $\psi_0 = e^{ikr \cos(\phi - \phi_0)}$ as

$$\Psi = \psi_0 + U(r), \quad U(r) = \sum_s u_s(r - r_s), \quad s = 0, \pm 1, \pm 2, \dots, \quad (1)$$

where the elementary wave scattered by the s th object (located at r_s) is given by an integral over its surface:

$$u_s(r - r_s) = \frac{1}{4i} \int [H_0(k | r - r_s' |) \partial_n \Psi(r_s'; \phi_0) - \partial_n H_0 \Psi_s] dA(r_s') \\ = \{H_0(k | r - r_s' |), \Psi(r_s', \phi_0)\}, \quad H_0 = H_0^{(1)}. \quad (2)$$

In addition, it is convenient to introduce the “multiple scattered amplitude” of cylinder 0 fixed at the origin:

$$G_0(\phi, \phi_0) = \{e^{-ikr \cos(\phi' - \phi)}, \Psi(r', \phi_0)\}, \quad y_0 = 0. \quad (3)$$

We are concerned only with an infinite number of scatterers ($N \rightarrow \infty$), but may keep N finite to facilitate discussion.

If we were to replace Ψ in (2) and (3) by the field on the isolated scatterer excited by ψ_0 , the present functions would reduce to their corresponding single scattered values. In particular, the single scattered form of (3) will be represented by g (which we assume to be known through the far field form of the corresponding scattered wave $e^{ikr}(2/i\pi kr)^{1/2}g$).

II. THE AVERAGE WAVE FUNCTION

General Form

The average wave function for a symmetrical distribution of identical scatterers along the y axis is given by

$$\langle U \rangle = \sum_s \int_{-\infty}^{\infty} \langle u \rangle_s \omega_s(y_s) dy_s, \quad (4)$$

$$\langle u \rangle_s = \langle \{H_0(kR_s), \Psi\} \rangle_s = \{H_0(kR_s), \langle \Psi \rangle_s\} \\ = \{H_0(kR_s), \langle \Psi \rangle_0\} e^{iky_s \sin \phi_0}, \quad R_s = |r - r' - y_s|. \quad (5)$$

$$\langle \Psi \rangle_0 = \psi_0 + \langle U \rangle_0 \\ = \psi_0 + \langle u \rangle_0 + \sum_{s \neq 0} \int \langle u(r - y_s) \rangle_0 \omega_0(y_0; y_s) dy_s. \quad (6)$$

[See reference 3 (15) to (23)]. Here $\langle u \rangle_s$ is the average elementary wave scattered by an object fixed at y_s , and $\langle \rangle_0$ is the analogous average when there is an additional scatterer fixed at y_0 . The function w_s is the one particle distribution function, and $w_0 dy_s$ is the conditional probability of finding s in the range dy_s once 0 is given in dy_0 . In (5) we used the symmetry which exists for an infinite number of identical scatterers and for an

infinite range of y_s (say $L \rightarrow \infty$) to write $\langle \Psi \rangle_s = e^{iky_s \sin \phi_0} \langle \Psi \rangle_0$; however, in order to facilitate the following discussion of the distribution functions, we may keep L finite.

The corresponding average scattering amplitude for an object fixed at $y_0 = 0$ is

$$G(\phi, \phi_0) \equiv \langle G_0(\phi, \phi_0) \rangle_0 = \{ e^{-ikr' \cos(\phi' - \phi)}, \langle \Psi \rangle_0 \}. \quad (7)$$

Our aim is to express G in terms of the known single scattered amplitude g , and in terms of the parameters of the distribution.

Distribution Functions

One-Particle Distribution Function: If we assume that all positions on the range available to one scatterer are equiprobable, then it may be shown* that

$$w_s(y_s) = \frac{e^{-|y_s - sB|/q}}{2q}; \quad \int_{-\infty}^{\infty} w_s dy_s = 1, \\ q = \frac{N(B - b)}{2} = \frac{ND}{2} = \int_{-\infty}^{\infty} |y_s - sB| w_s dy_s, \quad (8)$$

where B and b are the average and minimum separations of scatterer centers, and D is the "elbow room" per scatterer; $|y_s - sB|$ is the displacement of the scatterer from its mean position sB , and q is the average displacement from the mean. Note that w_s is essentially independent of s .

The limiting forms of w_s for the periodic and "rare gas" cases are identically those discussed previously.^{2,3} Thus if the average spacing reduces to the minimum ($B \rightarrow b$, $q \rightarrow 0$) then

$$w_s \rightarrow \delta(y_s - sB), \quad q \rightarrow 0; \quad (9)$$

i.e., in the periodic limit the function equals the Dirac delta function. On the other hand, if $B \gg b$, then

$$w_s \rightarrow \frac{1}{NB} = \frac{1}{L}; \quad (10)$$

thus in the rare gas limit, the function equals the reciprocal of the total range.³

We are concerned primarily with $N \rightarrow \infty$ and $L \rightarrow \infty$. For this case there are only two forms of w_s of direct interest: the periodic limit (9), and the more general form of (10) which we may write

$$w_s \rightarrow \frac{1}{N(B - b)}. \quad (10a)$$

This form holds for the "liquid state" or "dense gas," and reduces to (10) when the minimum separation is negligible compared to the average.

Distribution of Separations: We specify the distribution for the separation between any scatterer (say $s=0$) and its s th neighbor by

$$w_{0s} = \frac{(y_s - sb)^{s-1}}{D^s(s-1)!} e^{-(y_s - sb)/D}, \quad y_s > sb;$$

$$w_{0s} = 0, \quad y_s < sb.$$

$$\int_{-\infty}^{\infty} w_{0s} dy_s = \int_{sb}^{\infty} w_{0s} dy_s = 1. \quad (11)$$

This representation of w_{0s} was introduced by Zernike and Prins⁴ in their basic single-scattering treatment of the intensity pattern arising when liquids were irradiated with x -rays. (See second paper of reference 1 for a detailed derivation.)

Note that the pair distribution for two scatterers s and t equals $w_s w_t = P(s, t)$; our results are consistent in that for $N \rightarrow \infty$ we have $\int P(s, t) dy_s = w_t$ and $\int P(s, t) dy_t = w_s$ as required by fundamental considerations.

The total probability of finding any neighbor in a range $d\eta$ at some distance η from the center of a given scatterer is obtained by summing $w_{0s}(\eta)$ over all neighbors. A representation that is rapidly convergent near the periodic and gas limits is obtained by first taking the Laplace transform of $\sum w_{0s}$, and then its inverse.¹ This gives the sum of residues

$$w(\eta) = \sum_{s=1}^{\infty} w_{0s} = \sum_{\nu} \frac{e^{\eta \gamma_{\nu}}}{b + D e^{b \gamma_{\nu}}}, \quad (12)$$

where the γ_{ν} are the roots of $1 + D\gamma = e^{-\gamma b}$; the roots are of the form $\gamma_{\nu} = -\alpha_{|\nu|} \pm i\beta_{|\nu|}$ with the α 's and β 's real and positive.

The contribution of the pole at the origin, $\nu=0$, is simply $1/B$; i.e., the average number of cylinders in unit length of distribution. If $D/b = (B/b) - 1$ is very large, then $b\alpha_{|\nu|} \approx \Gamma + \frac{1}{2} \ln [1 + (2|\nu| - 1)^2 \pi^2 / \Gamma^2]$, and $b\beta_{|\nu|} \approx (2|\nu| - 1)\pi(1 + 1/\Gamma)$ with $\Gamma = \ln [(D/b) \ln (D/b)]$; retaining only the terms $\nu=0, \pm 1$, we write $\gamma_0=0$, $\gamma_{\pm 1} = -\alpha \pm i\beta$ and obtain

$$w(\eta) \approx [1 + e^{-\alpha(\eta-b)} 2 \cos \beta(\eta - b)]/B, \quad \eta > b \ll B. \quad (12a)$$

Hence w approaches $1/B$ in an exponentially decreasing oscillatory fashion as η increases. The distribution of neighbors is "ordered" only in the immediate vicinity of each cylinder (corresponding to the "local order" of liquid state theory). For very large D/b , we note that w is essentially a step function which equals zero for $\eta < b$ and $1/B$ thereafter. In the limit $D/b \rightarrow \infty$, (12a) reduces to the results for a "uniform" distribution; i.e., $w(\eta) = 1/B$ for all values of η .

On the other hand if $D/b \rightarrow 0$, then the roots may be approximated by

$$\gamma_{\nu} = \pm \frac{i2\nu\pi}{B} \left[1 + \frac{D^2(2\nu\pi)^2}{3B^3} \right] - \frac{2(\nu\pi D)^2}{B^3};$$

consequently (12) reduces to

$$w(\eta) \rightarrow \frac{1}{B} \sum e^{i2\pi\nu\eta/B} = \sum_{m=1}^{\infty} \delta(\eta - mb), \quad \eta > b. \quad (12b)$$

Thus in the limit $D \rightarrow 0$, we see that w reduces to the results for a diffraction grating of spacing b .

Average Wave Function

The Wave Function: Substituting w_s into $\langle U \rangle$ of (4) we may write

$$\langle U \rangle = \{ \mathfrak{G}, \langle \Psi \rangle_0 \},$$

$$\mathfrak{G} = \sum_{s=-\infty}^{\infty} \int_{-\infty}^{\infty} H_0(kR_s) e^{ik y_s \sin \phi_0} \frac{e^{-|y_s - sb|/q}}{2q} dy_s \quad (13)$$

where \mathfrak{G} is the Green's function for the reduced problem. Writing H_0 as an integral of plane waves, *i.e.*,

$$H_0(k | \mathbf{r} - \mathbf{r}' - \mathbf{y}_s |) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{ik|x-x'| \sqrt{1-t^2} + ik(y-y'-y_s)t}}{\sqrt{1-t^2}} dt \quad (14)$$

(where

$$\sqrt{1-t^2} = i\sqrt{t^2-1} \quad \text{for } t > 1),$$

yields

$$\mathfrak{G} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{ik|x-x'| \sqrt{1-t^2} + ik(y-y'-y_s)t}}{\sqrt{1-t^2}} I(t) dt, \quad (15)$$

$$I(t) = \sum_{s=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-ik(t-\sin \phi_0)y_s - |y_s - sb|/q} dy_s / 2q$$

$$= \frac{1}{2} \left[\frac{1}{1 + iqk(t - \sin \phi_0)} + \frac{1}{1 - iqk(t - \sin \phi_0)} \right]$$

$$\cdot \sum_{s=-\infty}^{\infty} e^{-isb(t - \sin \phi_0)}$$

$$= \left[\frac{1}{1 + q^2 k^2 (t - \sin \phi_0)^2} \right] \frac{2k}{kB} \cdot \sum_{\nu=-\infty}^{\infty} \delta \left(t - \sin \phi_0 - \frac{2\nu\pi}{kB} \right)$$

$$\nu = 0, \pm 1. \quad (16)$$

Thus, if $B = b(q=0)$, then I reduces to the periodic δ function. On the other hand, if $B \approx b$, then the term containing q^2 vanishes for $t = \sin \phi_0$, but is infinite otherwise; hence I equals $(2\pi/kB)\delta(t - \sin \phi_0)$ for all non-periodic cases.

Substituting the final form of I into \mathfrak{G} and evaluating the remaining integral gives

$$\mathfrak{G} = 2 \sum_{\nu} C_{\nu} \frac{e^{ik(y-y') \sin \phi_{\nu} + ik|x-x'| \cos \phi_{\nu}}}{1 + (q2\nu\pi/B)^2}$$

$$C_{\nu} = \frac{1}{kB \cos \phi_{\nu}}, \quad \sin \phi_{\nu} = \sin \phi_0 + \frac{2\nu\pi}{kB}. \quad (17)$$

Thus, for the periodic case, (17) becomes identical with the Green's function for the grating given in reference 2 (20). On the other hand, if $B \neq b$, then only the term $\nu=0$ of (17) remains and \mathfrak{G} reduces to the value for the "rare gas" given in reference 3 (20); *i.e.*, the gas and liquid give the identical result.

Substituting (17) into (13), and using the definition of the multiple scattered amplitude of (7), we obtain the forward scattered field

$$\langle U \rangle_> = 2 \sum \frac{C_{\nu} \psi_{\nu} G(\phi_{\nu}, \phi_0)}{1 + (q2\nu\pi/B)^2}, \quad \psi_{\nu} = e^{ikr \cos(\phi - \phi_{\nu})}, \quad (18)$$

and the reflected field

$$\langle U \rangle_< = 2 \sum \frac{C_{\nu} \psi_{\nu'} G(\pi - \phi_{\nu}, \phi_0)}{1 + (q2\nu\pi/B)^2}, \quad \psi_{\nu'} = e^{ikr \cos(\phi - \pi + \phi_{\nu})}; \quad (19)$$

where $\psi_{\nu'}$ is the image of ψ_{ν} in the plane of the distribution. If $q=0$, then (18) and (19) reduce to the results for the grating: the terms corresponding to $\sin \phi_{\nu} < 1$ are the usual propagating modes (or spectral orders), and those for $\sin \phi_{\nu} > 1$ are the evanescent modes. On the other hand, if $q \rightarrow \infty$, then only the "specular modes" ϕ_0 and $\phi_{0'}$ remain. (See references 2 and 3 for detailed discussion of these limiting cases.)

The preceding method was followed to stress the relation of the present general problem to the special cases treated previously;^{2,3} *i.e.*, we used essentially the previous procedure. However, we can arrive at (18) and (19) more directly by rewriting $\langle u \rangle_s$ of (5) in the form

$$\langle u \rangle_s = 2 \int_{-\infty}^{\infty} C_{\mu} G(\phi_{\mu}, \phi_0) e^{ikr \cos(\phi_{\mu} + \phi) + ik y_s (\sin \phi_0 - \sin \phi_{\mu})} d\mu,$$

$$\sin \phi_{\mu} = \sin \phi_0 + \frac{2\mu\pi}{kB}, \quad (5a)$$

which holds for $-\pi/2 < \phi < \pi/2$; similarly for the remaining two quadrants we replace ϕ_{μ} by $\pi - \phi_{\mu}$. This representation follows on substituting H_0 as in (14) (with $t = \sin \phi_{\mu}$, $dt = 2\pi d\mu/kB$) into (2) and using the definition of G as in (3). Thus (13) may be written

$$\langle U \rangle = \sum_s \int \langle u \rangle_s w_s dy_s$$

$$= \int_{-\infty}^{\infty} 2C_{\mu} G(\phi_{\mu}, \phi_0) \psi_{\mu} I(\mu) d\mu \quad (13a)$$

where $I(\mu)$ is the function of (16) with $t - \sin \phi_0 = 2\pi\mu/kB$. Substituting the final form of I into (13a) and evaluating the integral (the δ functions contributing when μ is an integer) gives (18) directly. Similarly (19) follows on replacing ϕ_{μ} by $\pi - \phi_{\mu}$.

The Multiple Scattered Amplitude: In order to obtain a more explicit form for the scattering amplitude G than (7), we must determine $\langle \Psi \rangle_0$ of (6). Proceeding more or less as for (13)ff, we obtain initially

$$\langle U \rangle_0 - \langle u \rangle_0 = \sum_{s \neq 0} \int \{ H_0(kR_s), \langle \Psi \rangle_{0s} \} w_{0s} dy_s. \quad (20)$$

Using the approximation $\langle \Psi \rangle_{0s} \approx \langle \Psi \rangle_s = e^{iky_s \sin \phi_0} \langle \Psi \rangle_0$ which is exact for the periodic limit, and plausible otherwise [see discussion of (24) in reference 3] we obtain

$$\langle U \rangle_0 - \langle u \rangle_0 = \{ \mathcal{C}, \langle \Psi \rangle_0 \}, \quad (21)$$

where \mathcal{H} is of the form of \mathcal{G} of (13) with $\sum_s w_s$ replaced by $\sum_{s \neq 0} w_{0s}$. Thus, we may write \mathcal{H} as (15) with the previous I replaced by

$$\begin{aligned} J &= \sum_{s \neq 0} \int e^{ik y_s (\sin \phi_0 - t)} w_{0s} dy_s \\ &= 2 \operatorname{Re} \sum_{s=1}^{\infty} \int_{sb}^{\infty} \frac{(y_s - sb)^{s-1}}{D^s (s-1)!} e^{ik y_s (\sin \phi_0 - t) - (y_s - sb)/D} dy_s \\ &= 2 \operatorname{Re} \sum_{s=1}^{\infty} \left[\frac{e^{ikb(\sin \phi_0 - t)}}{1 - ikD(\sin \phi_0 - t)} \right]^s. \end{aligned} \quad (22)$$

We may rewrite the final form of (22) as

$$\begin{aligned} J &= -\frac{\zeta^2}{\zeta^2 + d^2} + \frac{\epsilon(2 - \zeta^2)}{(1 - \epsilon)\zeta^2 + d^2 + \epsilon^2 \left(1 - \frac{\zeta^2}{2}\right)} \\ &= J_1 + J_2(\epsilon), \quad \epsilon \rightarrow 0; \\ d &= kD(\sin \phi_0 - t), \\ \zeta^2 &= 2(1 + d \sin \beta - \cos \beta) \\ \beta &= kb(\sin \phi_0 - t), \end{aligned} \quad (23)$$

and use

$$J_2 \rightarrow \frac{2\epsilon}{(\zeta^2 + d^2) + \epsilon^2} = 2\pi\delta(\sqrt{\zeta^2 + d^2})$$

to reduce (23) to

$$J = -1 + \frac{d^2}{\zeta^2 + d^2} + 2\pi\delta(\sqrt{\zeta^2 + d^2}). \quad (24)$$

In the periodic limit ($d \rightarrow 0$, $\zeta \rightarrow 2 \sin \beta/2$), J_1 reduces to -1 , and J_2 to the periodic δ function; *i.e.*, for $d=0$, the zeros of ζ occur for $\beta/2 = kb(\sin \phi_0 - t) = 2\nu\pi$. Thus,

$$J_{pd} = -1 + \frac{2\pi}{kb} \sum_{\nu} \delta(\sin \phi_{\nu} - t), \quad (25)$$

and consequently $\mathcal{H} = \mathcal{G}(I \leftrightarrow J)$ of (15) reduces to

$$\mathcal{H}_{pd} = -H_0 + 2 \sum C_{\nu} e^{ik(y-y') \sin \phi_{\nu} + ik|x-x'| \cos \phi_{\nu}}, \quad (26)$$

Substituting into (21), we see that the term in $-H_0$ cancels $\langle u \rangle_0$ in the expression $\langle U \rangle_0 = \langle u \rangle_0 + \{\mathcal{H}, \langle \Psi \rangle_0\}$ and we obtain

$$\langle U_0 \rangle \rightarrow \langle U \rangle = U \quad (27)$$

where $\langle U \rangle$ equals the function of (18) and (19) with $q=0$. Thus, the result for the periodic limit is identical with that of reference 2. More conveniently (as in reference 2), we use the plane-wave representation of H_0 [*i.e.*, (14) recast as for (5a)] in (26) and leave \mathcal{H} in the form

$$\begin{aligned} \mathcal{H}_{pd} &= \mathbf{S} 2C_{\nu} e^{ik(y-y') \sin \phi_{\nu} + ik|x-x'| \cos \phi_{\nu}}, \\ \mathbf{S} &= \left[\sum_{\nu=-\infty}^{\infty} - \int_{-\infty}^{\infty} d\nu \right] e^{ik\epsilon \cos \phi_{\nu}}, \quad \epsilon = |\epsilon| \rightarrow 0, \end{aligned} \quad (26a)$$

where ν is an integer for the sum operation, and a continuous variable for the integral. (The Abelian con-

vergence factor $e^{ik\epsilon \cos \phi_{\nu}} \rightarrow e^{-2\pi|\nu|\epsilon/b}$ is introduced to insure that the result of the operation exists.)

If $d \neq 0$, then J_2 yields only the single δ function at $\nu=0$, and $J_1 + 1 = d^2/(\zeta^2 + d^2)$ is essentially a "blurred" version of the periodic δ function. In particular, if $D/b \gg 1$, then

$$J_1 \rightarrow -2 [\sin kb(\sin \phi_0 - t)]/kB(\sin \phi_0 - t) \rightarrow 0,$$

and

$$J_{\tau} = \frac{2\pi}{kB} \delta(\sin \phi_0 - t) \quad (28)$$

for the rare gas. [Note that in the forward direction ($t = \sin \phi_0$) we get $J_1 \rightarrow -1 + (D/B)^2 = -2b/B + (b/B)^2 \rightarrow 0$.] Using (28) in \mathcal{H} gives only the term containing C_0 of (26). Consequently, for this case, (21) gives

$$\langle U \rangle_0 = \langle u \rangle_0 + \langle U \rangle \quad (29)$$

as obtained in reference 3.

We now reconsider the above in terms of the summed form w of (12), *i.e.*, instead of (22) we use

$$\begin{aligned} J &= 2 \operatorname{Re} \sum_{\nu} \frac{1}{b + De^{b\gamma_{\nu}}} \int_b^{\infty} e^{ik\eta(\sin \phi_0 - t) + \eta\gamma_{\nu}} d\eta \\ &= 2 \operatorname{Re} \sum \frac{e^{-iT_{\nu} - A_{\nu}}}{[1 + (D/b)e^{-A_{\nu} + iB_{\nu}}](iT_{\nu} + A_{\nu})} \end{aligned} \quad (22a)$$

where

$$\gamma_{\nu} = -\alpha_{\nu} + i\beta_{\nu} = -\frac{A_{\nu}}{b} + \frac{iB_{\nu}}{b},$$

$$T_{\nu} = kb(t - \sin \phi_0 - \beta_{\nu}/k).$$

In the periodic limit, $D \rightarrow 0$, we have $B_{\nu} = 2\nu\pi + \mathcal{O}(D^3)$, $T_{\nu} \rightarrow kb(t - \sin \phi_{\nu})$, and $A_{\nu} = \mathcal{O}(D^2)$. Thus,

$$\begin{aligned} J &\rightarrow 2 \operatorname{Re} \sum \frac{e^{-iT_{\nu}}}{iT_{\nu} + A_{\nu}} \\ &= 2 \operatorname{Re} \sum e^{-iT_{\nu}} \left[\frac{A_{\nu}}{T_{\nu}^2 + A_{\nu}^2} - \frac{iT_{\nu}}{T_{\nu}^2 + A_{\nu}^2} \right]. \end{aligned}$$

The first term in the brackets gives the periodic δ function of (25):

$$\begin{aligned} 2 \operatorname{Re} \sum e^{-iT_{\nu}} \frac{A_{\nu}}{T_{\nu}^2 + A_{\nu}^2} \\ = 2\pi \sum \cos T_{\nu} \delta(T_{\nu}) = \frac{2\pi}{kb} \sum \delta(\sin \phi_{\nu} - t). \end{aligned}$$

The second term equals⁵

$$2 \operatorname{Re} \sum_{x \rightarrow 1} \frac{e^{-i(T_0 - 2\nu\pi x)}}{i(T_0 - 2\nu\pi x)} = 2 \operatorname{Re} \sum_{x \rightarrow 1} \frac{e^{-iT_0/2x}}{2ix \sin(T_0/2x)} = -1$$

as in (25).

⁵ W. Magnus and F. Oberhettinger, "Formeln und Satze," Springer-Verlag, Berlin, Ger., p. 214; 1948. Here we used the limit of the first form instead of the result for $x=1$, because the limiting procedure is called for by the corresponding specialization of (22a).

In the gas limit, we keep only the term $\nu=0$ ($\gamma_0=0$) of (22a). Thus,

$$J \rightarrow 2 \operatorname{Re} \frac{e^{-iT_0}}{[1 + (D/b)](iT_0 + A)} \\ = \frac{2\pi}{kB} \delta(\sin \phi_0 - t) - \frac{2 \sin [kb(t - \sin \phi_0)]}{kB(t - \sin \phi_0)},$$

where the first term equals (28), and the remainder is the correction stated directly before (28).

Rare gas: Having obtained (29), it facilitates discussion to recall the subsequent steps of reference 3 leading to G . Thus, representing $\langle U \rangle$ in (29) as the mean of the transmitted and reflected forms $\langle U \rangle_>$ and $\langle U \rangle_<$ gives

$$\begin{aligned} \langle \Psi \rangle_0 &= \psi_0 + \langle U \rangle_0 = \psi_0 + \langle U \rangle + \langle u \rangle_0 \\ &= \psi_0 [1 + C_0 G(\phi_0, \phi_0)] \\ &\quad + \psi_0 C_0 G(\pi - \phi_0, \phi_0) + \langle u \rangle_0, \end{aligned} \quad (30)$$

which is of the form of the solution of the problem of an isolated scatterer excited by two plane waves (ψ_0 and its image ψ_0'). Consequently, from the superposition principle, it follows that the multiple scattered amplitude G of $\langle u \rangle_0$ is simply a linear combination of appropriate single scattered amplitudes g . Thus, by inspection of (30), we have

$$\begin{aligned} G(\phi, \phi_0) &= g(\phi, \phi_0) [1 + C_0 G(\phi_0, \phi_0)] \\ &\quad + g(\phi, \pi - \phi_0) C_0 G(\pi - \phi_0, \phi_0), \end{aligned} \quad (31)$$

which, on substituting first $\phi = \phi_0$ and then $\phi = \pi - \phi_0$, gives two simultaneous equations for the unknown values of G . Equivalently, on introducing the symmetry components

$$\begin{aligned} f_{\pm}(\phi, \pi - \phi_0) &= g(\phi, \pi - \phi_0) \pm g(\phi, \phi_0), \\ F_{\pm}(\phi, \pi - \phi_0) &= G(\phi, \pi - \phi_0) \pm G(\phi, \phi_0) \end{aligned} \quad (32)$$

we reduce (31) to the simpler form

$$F(\phi, \pi - \phi_0) = f(\phi, \pi - \phi_0) [1 + CF(\pi - \phi_0, \phi_0)]. \quad (33)$$

Letting $\phi = \phi_0$, solving for $F(\pi - \phi_0, \phi_0)$, and substituting back into (33) thus gives the explicit result

$$F(\phi, \pi - \phi_0) = \frac{f(\phi, \pi - \phi_0)}{1 - C_0 f(\phi_0, \pi - \phi_0)}. \quad (34)$$

[See reference 3 (25)ff].

Periodic Case: Similarly, for the periodic case, corresponding to (30), we obtain

$$\begin{aligned} \langle \Psi \rangle_0 &= \psi_0 + [\langle U \rangle_0 - \langle u \rangle_0] + \langle u \rangle_0 \\ &= \psi + \mathbf{S} C_{\nu} [\psi_{\nu} G(\phi_{\nu}, \phi_0) + \psi_{\nu}' G(\pi - \phi_{\nu}, \phi_0)] \\ &\quad + \langle u \rangle_0, \end{aligned} \quad (35)$$

where $\langle U \rangle_0 - \langle u \rangle_0 = U - u_0$ followed from (21) in terms of \mathcal{H} of (26a). Using the superposition principle, as previously, gives

$$\begin{aligned} G(\phi, \phi_0) &= g(\phi, \phi_0) \\ &\quad + \mathbf{S} C_{\nu} [g(\phi, \phi_{\nu}) G(\phi_{\nu}, \phi_0) + g(\phi, \pi - \phi_{\nu}) G(\pi - \phi_{\nu}, \phi_0)], \end{aligned} \quad (36)$$

or equivalently,

$$F_{\mu 0} = f_{\mu 0} + \mathbf{S} C_{\nu} f_{\mu \nu} F_{\nu 0}, \quad f_{\nu \mu} = f(\phi_{\nu}, \pi - \phi_{\mu}), \text{ etc.} \quad (37)$$

(See reference 2 and more recent papers⁶ for detailed discussions of the periodic case.)

General Case: For the general case, we use the alternative form of $\langle u \rangle_0$ as in (5a) to rewrite (21) in the form

$$\langle U \rangle_0 - \langle u \rangle_0 = \int_{-\infty}^{\infty} 2C_{\mu} G(\phi_{\mu}, \phi_0) \psi_{\mu} J(\mu) d\mu \equiv \mathfrak{U}, \quad (21a)$$

where J is the function of (22)ff and (22a)ff, with $t - \sin \phi = 2\pi\mu/kb$. With this plane-wave representation for $\langle U \rangle_0 - \langle u \rangle_0$, we write the average total field at a fixed scatterer as a collection of plane waves plus one outgoing cylindrical wave:

$$\langle \Psi \rangle_0 = \psi_0 + \langle U \rangle_0 = \psi_0 + \mathfrak{U} + \langle u \rangle_0, \quad (38)$$

where, as previously, we use the mean of the transmitted and reflected forms for \mathfrak{U} . The previous superposition procedure thus leads directly to the forms (36) for G , and (37) for the symmetry components F , with the operator \mathbf{S} now equalling

$$\mathbf{S} = \int_{-\infty}^{\infty} d\mu J(\mu) e^{ik|\epsilon| \cos \phi_{\mu}}, \quad \epsilon \rightarrow 0. \quad (39)$$

In the periodic limit, we use J of (25) and obtain \mathbf{S} of (26) directly. In the rare gas limit, we use J of (28), i.e., $2\pi\delta(\mu)/kb$.

The present result for G in terms of \mathbf{S} of (39) makes clear how all the special results obtained previously for arbitrary scatterers^{2,3} follow from the general "liquid" distribution. In addition, we also see that various special procedures developed for separable problems may be carried over directly. Thus, the same algebraic relations between the multiple scattered "coefficients" (the amplitudes of a Fourier series expansion of G in terms of ϕ) and their single scattered values given for the periodic distribution of circular⁶ and elliptic⁷ cylinders apply for the general distribution, provided that the previous functions \mathcal{H} are now defined in terms of the new operator (39).

Discounting the periodic limit, we use

$$J = J_1 + J_2 = J_1 + \delta(\mu), \quad (40)$$

as follows from (23)ff for a nonperiodic distribution. Substituting (40) into (39) and then carrying out the δ function integration in the form (37) gives

⁶ V. Twersky, "Notes on Scattering by Gratings," and "On the Scattering of Waves by the Infinite Grating of Circular Cylinders," Sylvania Electronic Defense Lab., Mountain View, Calif., Repts. EDL-M105 and EDL-E28, respectively; May, 1957, March, 1958. Reference 6 as used in the text applies only to the second of these.
⁷ J. E. Burke and V. Twersky, "On scattering of waves by the infinite grating of elliptic cylinders," presented at URSI meeting, Washington, D. C.; May, 1959.

$$F_{\mu 0} = f_{\mu 0}(1 + C_0 F_{00}) + \mathbf{S}_1 C_{\nu} f_{\mu \nu} F_{\nu 0} \quad (41)$$

where the subscript on the operator indicates that J is replaced by J_1 in (39). Letting $\mu=0$ in (41), we "solve" for F_{00} and then substitute back into (41) to obtain

$$F_{\mu 0} = p_{\mu 0} + \mathbf{S}_1 C_{\nu} p_{\mu \nu} F_{\nu 0}, \quad (42)$$

$$p_{\mu \nu} = f_{\mu \nu} + \frac{f_{\mu 0} C_{0 \nu}}{1 - C_{00}}, \quad C_0 = \frac{1}{k B \cos \phi_0}, \quad (43)$$

where $p_{\mu \nu}$, which is specified solely in terms of the known single scattered amplitude, is of the form of the "transformed amplitude" introduced for the periodic case² to make explicit the "mode coupling" associated with the grating anomalies of Wood and Strong.² (There we "suppressed" the near grazing mode, where as now we removed the zeroth or specular mode from the functional equation.)

In particular, the specular value F_{00} (the only explicit value required for the average wave function) equals

$$F_{00} = p_{00} + \mathbf{S}_1 C_{\nu} p_{0 \nu} F_{\nu 0} = \frac{f_{00} + \mathbf{S}_1 C_{\nu} f_{0 \nu} F_{\nu 0}}{1 - C_{00}}. \quad (44)$$

As shown previously,^{2,3} the symmetry components F are also the scattering amplitudes corresponding to a distribution of protuberances on a ground plane (which serves as a model for a large class of rough surface reflection phenomena), and the f 's are the analogous functions for an isolated protuberance. In a subsequent section, we use the known³ dependence of f on angles near grazing, and the form (44), to determine the angular dependence of the reflection coefficient near grazing.

III. THE AVERAGE INTENSITY

The intensity associated with a single configuration is the absolute square of Ψ of (1):

$$|\Psi|^2 = 1 + 2 \operatorname{Re} \psi_0^* U + |U|^2, \quad U = \sum u_s, \\ |U|^2 = \sum u_i \sum u_s^* = \sum |u_s|^2 + \sum_{s \neq i} \sum u_i u_s^*. \quad (45)$$

Its average value,

$$\langle |\Psi|^2 \rangle = 1 + 2 \operatorname{Re} \psi_0^* \langle U \rangle + \langle |U|^2 \rangle, \quad (46)$$

depends on $\langle U \rangle$ (the function determined in the previous section) and on the average scattered intensity

$$\langle |U|^2 \rangle = \sum_s \langle |u_s|^2 \rangle + \sum_{s \neq i} \sum_i \langle u_i u_s^* \rangle \\ = \sum \int \langle |u_s|^2 \rangle \omega_s dy_s \\ + \sum' \sum \iint \langle u_i u_s^* \rangle w_s w_{s i} dy_s dy_i, \quad (47)$$

where the prime on the double sum means $s \neq i$, and where $w_s w_{s i}$ is the two particle distribution function. Here the term involving the single sum is the "incoher-

ent scattering," and the other is essentially the "coherent scattering."

We use the approximations³ $\langle u_s u_i^* \rangle_{s i} \approx \langle u_s \rangle_s \langle u_i^* \rangle_i$, which essentially neglects contributions to the excitation of a scatterer arising from the fluctuations of the average radiation scattered by the others. Then in terms of $\langle u \rangle_s$ of (5a), the function $\langle |U|^2 \rangle$ of (47) reduces to

$$\langle |U|^2 \rangle = \int_{-\infty}^{\infty} \frac{d(\sin \tau)}{\pi \cos \tau} e^{ik\tau \cos(\tau - \phi)} G(\tau, \phi_0) \\ \cdot \int_{-\infty}^{\infty} \frac{d(\sin \tau')}{\pi \cos \tau'} e^{-ik\tau' \cos(\tau' - \phi)} G^*(\tau', \phi_0) M(\tau, \tau'). \quad (48)$$

$$M = \sum \int_{-\infty}^{\infty} e^{iky_s(\sin \tau - \sin \tau')} w_s dy_s \\ + \sum' \sum \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{iky_s(\sin \phi_0 - \sin \tau) - ik y_i(\sin \phi_0 - \sin \tau')} \\ w_s w_{s i} dy_s dy_i \\ = \left[1 + \sum' \int dy_i e^{ik(y - y_s)(\sin \phi_0 - \sin \tau)} w_{s i} \right] \\ \cdot \sum \int_{-\infty}^{\infty} e^{iky_s(\sin \tau - \sin \tau')} w_s dy_s \\ = [1 + J(\tau, \phi_0)] \cdot I(\tau, \tau'), \quad (49)$$

where $I(\tau, \tau')$ is the function of (16) (with $\sin \tau - \sin \tau'$ corresponding to the previous $t - \sin \phi_0$), and J is the function of (22).

Thus, in the periodic limit we have, from (16) and (25),

$$M = \left(\frac{2\pi}{kb} \right)^2 \sum \delta(\sin \phi_0 - \sin \tau) \\ \cdot \sum \delta \left(\sin \tau - \sin \tau' - \frac{2\nu\pi}{kb} \right). \quad (50)$$

Substituting (50) into (48) and evaluating the integrals gives simply the square of (18) and (19) as required.

On the other hand, in the rare gas limit, we obtain only the $\mu = \nu = 0$ terms of (50),

$$M = \left(\frac{2\pi}{kb} \right)^2 \delta(\sin \phi_0 - \sin \tau) \delta(\sin \tau - \sin \tau'), \quad (51)$$

and (48) reduces to

$$\langle |U|^2 \rangle = \frac{2}{\pi k B} \int \left| \frac{G(\tau, \theta_0)}{\cos \tau} \right|^2 d(\sin \tau) + \langle |U \rangle|^2 \\ = V + \langle |U \rangle|^2, \quad (52)$$

where V is the variance of U . Reworking the y_s integration and $|H_0|^2$ back into V , and using the asymptotic form $H_0 = (2/i\pi k r)^{1/2} e^{ikr}$ gives the far-field form

$$V \sim \frac{2}{\pi k B} \int \frac{|G(\phi_s, \phi_0)|^2}{|r - y_s|} dy_s, \quad \tan \phi_s = \frac{y - y_s}{x}, \quad (53)$$

as in (43) of reference 3.

Thus, we have shown that both the average wave function and intensity of the general one-dimensional "liquid distribution" specified by w_s of (8) and w_{0s} of (11) go over completely in the appropriate limits to the results obtained previously for the periodic² and rare gas³ cases.

More generally, we evaluate the integral over τ' directly to obtain a function of the form $\langle U \rangle^*$ treated in the previous section, *i.e.*,

$$\begin{aligned} & \int_{-\infty}^{\infty} \frac{d(\sin \tau')}{\pi \cos \tau'} e^{-ikr \cos(\tau' - \phi)} G^*(\tau', \phi_0) \\ & \cdot \sum_{\nu} \frac{2\pi}{kB} \frac{\delta(\sin \tau - \sin \tau' - 2\nu\pi/kB)}{[1 + q^2 k^2 (\sin \tau - \sin \tau')^2]} \\ & = \sum_{\nu} \frac{2\pi e^{-ikr \cos(\tau - \phi)} G^*(\tau, \phi_0)}{kB \cos \tau [1 + (2\nu\pi q/B)^2]}, \\ & \sin \tau_{\nu} = \sin \tau + \frac{2\nu\pi}{kB} \end{aligned} \quad (54)$$

Except for the periodic limit (which we henceforth neglect) this expression reduces to the single term for $\nu=0$, and consequently (48) equals

$$\langle |U|^2 \rangle = \frac{2}{\pi kB} \int_{-\infty}^{\infty} \left| \frac{G(\tau, \phi_0)}{\cos \tau} \right|^2 [1 + J(\tau, \phi_0)] d(\sin \tau) \quad (55)$$

which, of course, reduces to (53) in the gas limit. Writing $J = J_1 + J_2 = J_1 + (2\pi/kB) \delta(\sin \phi_0 - \sin \tau)$ [as follows from (23)ff for the nonperiodic limit] and transposing the δ function contribution $|\langle U \rangle|^2$, gives

$$\begin{aligned} V &= \langle |U|^2 \rangle - |\langle U \rangle|^2 \\ &= \frac{2}{\pi kB} \int_{-\infty}^{\infty} \left| \frac{G(\tau, \phi_0)}{\cos \tau} \right|^2 (1 + J_1) d(\sin \tau) \\ &\sim \frac{2}{\pi kB} \int \frac{|G(\phi, \phi_0)|^2}{|r - y_s|} (1 + J_1) dy_s, \end{aligned} \quad (56)$$

where the final representation is the far-field form analogous to (53).

IV. THE AVERAGE ENERGY FLUX

Proceeding essentially as in reference 3 (44)ff, we write the ensemble average of the total time-averaged energy flux per unit area divided by the time-averaged incident flux density as

$$\begin{aligned} \langle S \rangle &= \text{Re} \langle \Psi^* \nabla \Psi / ik \rangle \\ &= \text{Re} \left[\mathbf{o} + \psi_0^* (\mathbf{o} + \nabla / ik) \sum \langle u_s \rangle \right. \\ &\quad \left. + \sum \sum \langle u_i^* \nabla u_s / ik \rangle \right], \end{aligned} \quad (57)$$

where \mathbf{o} is a unit vector in the direction of incidence. Neglecting the periodic limit (since the previous section demonstrates that we would simply obtain the results treated in detail in reference 2), we obtain

$$\begin{aligned} \langle S_{\rangle} &= |\psi_0 + \langle U_{\rangle}|^2 \mathbf{o} + I_{\rangle}, \\ \langle S_{\langle} &= \mathbf{o} + \text{Re}(\psi_0^* \langle U_{\langle})(\mathbf{o} + \mathbf{o}') + \langle |U_{\langle}|^2 \mathbf{o} + I_{\langle}, \\ \left. \begin{matrix} \mathbf{o} \\ \mathbf{o}' \end{matrix} \right\} &= \pm i \cos \phi_0 + j \sin \phi_0, \end{aligned} \quad (58)$$

where, as for (18) and (19), the subscripts \rangle and \langle indicated transmitted and reflected, and where the incoherent flux equals

$$\begin{aligned} I &= V[s] \sim \frac{2}{\pi kB} \int \frac{|G(\phi_s, \phi_0)|^2}{|r - y_s|} (1 + J_1) s dy_s, \\ s &= i \cos \phi_s + j \sin \phi_s, \end{aligned} \quad (59)$$

i.e., I , the "sum" of the incoherent fluxes of all scatterers differs from V of (56) in that its kernel possesses a unit vector pointing from the scatterer at y_s to the field point r .

Proceeding as in reference 3, we rewrite (59) as

$$\begin{aligned} I &= \int \sigma(\phi_s, \phi_0) \frac{s}{\cos \phi_s} d\phi_s, \\ \sigma &= \frac{2}{\pi kB} |G(\phi_s, \phi_0)|^2 (1 + J_1), \end{aligned} \quad (60)$$

where σ is the differential scattering cross section per unit area of distribution. (We may introduce antenna gain patterns, etc., in (60) to obtain practical forms.)

Since (58) is identical in form with (48),³ the energy theorem derived previously for the rare gas of lossless scatterers also applies for the present more general distribution. Thus, equating the normal components of $\langle S_{\rangle}$ and $\langle S_{\langle}$ gives

$$\begin{aligned} 1 &= T + R + P/B \cos \phi_0, \\ T &= |\psi_0 + \langle U_{\rangle}|^2 = |1 + 2C_0 G(\phi_0, \phi_0)|^2, \\ R &= |\langle U_{\langle}|^2 = |2C_0 G(\pi - \phi_0, \phi_0)|^2, \\ \frac{P}{B} &= \int_0^{2\pi} \sigma(\phi, \phi_0) d\phi; \end{aligned} \quad (61)$$

where T and R are the transmission and reflection coefficients, and P is the total scattering cross section of one object in the distribution. Since $1/B$ is the average number of scatterers in unit length, and $1/B \cos \phi_0$ is the average number of scatterers "illuminated" by unit area of incident wave, we may interpret (61) as follows: the average power transmitted, reflected, and incoherently scattered by the area of the distribution illuminated by unit area of incident wave is equal to the incident power density. The above may be directly extended to absorbing scatterers by defining P as the sum of the present scattering cross section plus a surface integral corresponding to the average absorption cross section.

Rewriting the theorem of (61) in terms of G gives

$$-\operatorname{Re} G(\phi_0, \phi_0) = \frac{1}{2\pi} \int_0^{2\pi} |G(\phi, \phi_0)|^2 [1 + J_1] d\phi \\ + C_0 [|G(\phi, \phi_0)|^2 + |G(\pi - \phi_0, \phi_0)|^2] \quad (62)$$

as compared with the analogous theorem for the single scattered amplitude

$$-\operatorname{Re} g(\phi_0, \phi_0) = \frac{1}{2\pi} \int_0^{2\pi} |g(\phi, \phi_0)|^2 d\phi. \quad (63)$$

For the rare gas limit, J_1 equals zero and (62) reduces to (54);³ for the periodic limit, we restore the additional δ function terms of (50) in (58), set $1 + J_1$ equal to zero, and obtain (36)², *i.e.*,

$$-\operatorname{Re} G(\phi_0, \phi_0) \\ = \sum_p C_p [|G(\phi_p, \phi_0)|^2 + |G(\pi - \phi_p, \phi_0)|^2] \quad (64)$$

where the subscript p indicates that sum is only over the propagating modes. In (63), the energy "lost" from the incident wave through interference with the forward scattered wave appears as radiation in all other directions; in (64), it appears in the discrete directions corresponding to the usual spectral orders of the grating; in the rare gas, it is a superposition of radiation over the continuum of directions plus the discrete specular contributions (*i.e.*, transmitted and reflected); in (62) the additional term J_1 "modulates" the continuum term and peaks it at discrete angles to result in a blurred version of (64).

Similarly, we obtain analogous results for the symmetry components F , and for the corresponding intensities for distributions of protuberances on a ground plane. Thus

$$1 = R + P/B \cos \phi_0, \\ R = |1 + 2C_0 F(\phi_0, \pi - \phi_0)|^2, \quad (65) \\ \frac{P}{B} = \int_{-\pi/2}^{\pi/2} \sigma(\phi, \pi - \phi_0) d\phi = \frac{2}{\pi k B} \int |F(\phi, \pi - \phi_0)|^2 d\phi \\ -\operatorname{Re} F_0(\phi_0, \pi - \phi_0) = \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} |F(\phi, \pi - \phi_0)|^2 (1 + J_1) d\phi \\ + \frac{|F(\phi_0, \pi - \phi_0)|^2}{k B \cos \phi_0}. \quad (66)$$

V. APPLICATIONS

Single Scattering

The previous multiple scattering treatment of the periodic³ and rare gas cases² indicate that the effects of multiple scattering are primarily significant when there is a scattered mode near grazing, or the direction of incidence is near grazing, or the scatterers are relatively closely packed; *i.e.*, when the average separation is of the order of their size. Discounting these ranges of the parameters temporarily we may then, in general, replace G by its single scattered value g .

It was essentially the single scattered form of the incoherent intensity which was considered by Zernike and Prins.⁴ They treated a finite range of distribution (small compared to the distance of observation) and obtained essentially

$$V_1 = \left(\frac{2}{\pi k r}\right) \frac{L}{B} |g(\phi, \phi_0)|^2 (1 + J_1) \\ = \left[\left(\frac{2}{\pi k r}\right) |g(\phi, \phi_0)|^2\right] N \left[\frac{d^2}{\xi^2 + d^2}\right], \quad (67)$$

$$d = k D (\sin \phi_0 - \sin \phi),$$

$$\xi^2 = 2(1 + d \sin \beta - \cos \beta),$$

$$\beta = k b (\sin \phi_0 - \sin \phi).$$

Here $(2/\pi k r) |g|^2$ is the far-field scattered intensity for an isolated scatterer, $N = L/B$ is the number of scatterers, and $1 + J_1$ is the "modulation" introduced by the distribution. See their paper⁴ for plots of $1 + J_1$ as a function of d for several values of D/b and for corresponding plots of $w = \sum w_{0s}(\eta)$ of (11) as a function of η . Their curves for $D/b = (B - b)/b = \frac{1}{2}$ (where B and b are the average and minimum spacings respectively) show that w is practically constant in the range $\eta > 6b$, and the corresponding plot of $1 + J_1$ is a slowly varying function of d ; on the other hand for $D/b = 1/10$, both curves show pronounced (initially almost equally spaced) maxima whose heights decrease with increasing argument. These two sets, plus their remaining set for $D/b = \frac{1}{4}$, show clearly the dependence on D/B in the transition from the "more random" to the "more ordered" ranges. See their paper for details.⁴

Note, however, that their results near the forward direction do not describe the total scattering for a finite number of scatterers. The discrepancy arises from the fact that (67) which was derived for $N = \infty$ does not include the coherent scattering. To restore the coherent effects, we rework the derivation for the case of finite N ; however, for simplicity, and to avoid redefining w_s and w_{st} , we keep the range of the distribution infinite. We proceed essentially as in reference 1.

Using $|r - y_s| \sim r - y_s \sin \phi$, we write the single scattered field of a cylinder at 0, y_s as

$$u_s(|r - y_s|) = \sqrt{\frac{2}{\pi k r}} e^{ikr} g(\phi, \phi_0) e^{-ik y_s (\sin \phi - \sin \phi_0)} \quad (68)$$

and the total single scattered intensity as

$$\langle |U_1|^2 \rangle = \left[\frac{2}{\pi k r} |g(\phi, \phi_0)|^2\right] M(\phi, \phi_0), \\ M = M_1 + M_2, \quad M_1 = \sum_{t=-\infty}^{\infty} \int_{-\infty}^{\infty} w_t dy_t \\ M_2 = \sum_t \sum'_s \iint e^{i(y_s - y_t)k(\sin \phi - \sin \phi_0)} w_t w_{ts} dy_t dy_s. \quad (69)$$

[See (49).] With (8), we reduce M_1 to

$$M_1 = (2n + 1) = N, \quad (70)$$

where N is the total number of scatterers. To evaluate M_2 , we change the variable y_s to $\eta = y_s - y_t$; since w_{ts} is a function of $y_s - y_t$, the y_t integral again gives unity and we obtain simply

$$M_2 = \sum_{t=-n}^n \int_{sb}^{\infty} \left[\sum_{s=1}^{n+t} e^{iK\eta} + \sum_{s=1}^{n-t} e^{-iK\eta} \right] w_{0s}(\eta) d\eta, \quad (71)$$

where $K = k(\sin \phi - \sin \phi_0)$, and w_{0s} is given in (11). Evaluating the integral over η [as in (22)], and summing over s and t gives

$$\begin{aligned} M_2 &= 2 \operatorname{Re} \sum_{t=-n}^n \sum_{s=1}^{n+t} Q^s \\ &= 2 \operatorname{Re} \left\{ \frac{Q}{1-Q} \left[N - \frac{1-Q^N}{1-Q} \right] \right\}, \\ Q &= \frac{e^{iKb}}{1-iKD}, \end{aligned} \quad (72)$$

where $Q/(1-Q)$ is the previous result J_1 of (23) as obtained for infinite N .

If D/b is moderately large and $\phi \neq \phi_0$, $\pi - \phi_0$ (i.e., $K \neq 0$), then we need retain only the term proportional to N . For this case, $M_1 + M_2$ gives simply $N(1 + J_1)$ and (69) reduces to V_1 of (67) as obtained by Zernike and Prins.⁴ On the other hand, if $K \rightarrow 0$, then all terms of M_2 must be retained. For this case,

$$\frac{Q}{1-Q} \rightarrow \frac{i - Kb}{KB}, \quad \frac{1-Q^N}{1-Q} \rightarrow N + \frac{iKBN(N-1)}{2}$$

and M_2 reduces to $N(N-1)$; consequently

$$M = M_1 + M_2 \rightarrow N^2 \text{ as } K \rightarrow 0, \quad (73)$$

as required by elementary physical considerations.

More generally, the limiting cases for finite N can be obtained only by retaining the "end corrections" in M_2 . Thus, near the rare gas case $B/b \gg 1$, we have

$$\begin{aligned} 2 \operatorname{Re} \frac{Q}{1-Q} &\approx \frac{-2 \sin bK}{BK}, \\ -2 \operatorname{Re} \frac{Q(1-Q^N)}{(1-Q)^2} &\approx N(N-1) \frac{\sin^2 (NBK/2)}{(NBK/2)^2}, \end{aligned}$$

and consequently

$$\begin{aligned} M &= N \left[1 + (N-1) \frac{\sin^2 (NBK/2)}{(NBK/2)^2} - \frac{2 \sin bK}{BK} \right], \\ K &= k(\sin \phi - \sin \phi_0), \end{aligned} \quad (74)$$

where the second term is essentially the usual Fraunhofer result for an aperture of width NB . The second

term is the significant one at the specular angles $K=0$ ($\phi = \phi_0, \pi - \phi_0$); the second and third are significant near these angles, and only the first is required for large values of K . Since the second term drops off much more rapidly than the third as K increases, the third is significant over a larger range of $\phi - \phi_0$. Thus, for $|\phi - \phi_0| > \lambda/NB$ (the first zero of the second term), the first and third yield a series of maxima whose magnitudes hinge on $2b/B$. The pronounced specular lobes will be flanked by two fairly distinct sidelobes.

On the other hand, if $D/b \rightarrow 0$, then $Q \rightarrow e^{i(b+D)K} = e^{iBK}$,

$$\begin{aligned} \frac{Q}{1-Q} &\rightarrow \frac{ie^{iBK/2}}{2 \sin (BK/2)}, \\ \frac{1-Q^N}{1-Q} &\rightarrow e^{iBK(N-1)/2} \frac{\sin (KBN/2)}{\sin (KB/2)}. \end{aligned}$$

Consequently

$$M = \frac{\sin^2 (NBK/2)}{\sin^2 (BK/2)} \quad (75)$$

where is the usual Fraunhofer result for a finite grating of spacing B .

Scatterers in the Far-Field of Each Other

The first paper of reference 1 treated a "liquid distribution" of circular cylinders by means of separation of variables. We obtained closed forms for the monopole and dipole multiple scattered coefficients in terms of their single scattered coefficients, as well as a heuristic closed form for the complete multiple scattering amplitude in terms of its single scattered value. We consider first the scattering amplitude, and reserve discussion of the coefficients for a following section.

By treating each scatterer as if it were in the far field of all neighbors (or equivalently, by using the asymptotic form of (2) for $k|\mathbf{r} - \mathbf{r}_s| \gg 1$) we found¹

$$\begin{aligned} G(\phi, \phi_0) &= g(\phi, \phi_0) + g\left(\phi, \frac{\pi}{2}\right) G\left(\frac{\pi}{2}, \phi_0\right) \mathcal{C}_- \\ &\quad + g\left(\phi, -\frac{\pi}{2}\right) G\left(-\frac{\pi}{2}, \phi_0\right) \mathcal{C}_+ \\ \mathcal{C}_{\pm} &= \sqrt{\frac{2}{i\pi}} \int_0^{\infty} \frac{e^{ik\eta(1 \mp \sin \phi_0)}}{\sqrt{kn}} w(\eta) d\eta \end{aligned} \quad (76)$$

[see reference 1 (28) and (33)], where $w(\eta) = \sum w_{0s}$ of (11) or (12). Substituting first $\phi = \pi/2$ and then $\phi = -\pi/2$ gives two simultaneous equations which may be solved directly for the two special values of G on the right. [See reference 1 (30).]

The corresponding symmetry components, as defined in (32), were found to equal

$$F_+(\phi, \pi - \phi_0)$$

$$= f_+(\phi, \pi - \phi_0) + \frac{1}{2} \left[f\left(\phi, \frac{\pi}{2}\right) F_+\left(\frac{\pi}{2}, \phi_0\right) \mathcal{R}_- \right. \\ \left. + f\left(\phi, -\frac{\pi}{2}\right) G\left(-\frac{\pi}{2}, \phi_0\right) \mathcal{R}_+ \right] \\ = g(\phi, \phi_0) + g(\phi, \pi - \phi_0) + 2 \left[g\left(\phi, \frac{\pi}{2}\right) G\left(\frac{\pi}{2}, \phi_0\right) \mathcal{R}_- \right. \\ \left. + g\left(\phi, -\frac{\pi}{2}\right) G\left(-\frac{\pi}{2}, \phi_0\right) \mathcal{R}_+ \right],$$

$$F_-(\phi, \pi - \phi_0) = f_-(\phi, \pi - \phi_0) \\ = g(\phi, \phi_0) - g(\phi, \pi - \phi_0). \quad (77)$$

To this approximation, the antisymmetric amplitude (or, equivalently, the amplitude for protuberances on a ground plane excited by horizontal polarization) shows no effects of multiple scattering; *i.e.*, because

$$f_-\left(\phi, \frac{\pi}{2}\right) \\ = \left[g\left(\phi, \frac{\pi}{2} - \epsilon\right) - g\left(\phi, \pi - \left[\frac{\pi}{2} - \epsilon\right]\right) \right]_{\epsilon \rightarrow 0} \rightarrow 0.$$

The functions \mathcal{R}_\pm were expressed in terms of the Fresnel integrals. Thus [as in reference 1(33)ff and (42)] we found

$$\mathcal{R}_\pm = 2 \sum_{\nu=-\infty}^{\infty} \frac{\mathcal{F}(NL_\nu) - \mathcal{F}(L_\nu)}{[1 + (D/b)e^{b\gamma_\nu}] \sqrt{ikbL_\nu}}, \\ (L_\nu)_\pm = kb(1 \mp \sin \phi_0) - ib\gamma_\nu, \quad (78)$$

where \mathcal{F} is the Fresnel integral, such that

$$\mathcal{F}(\eta) \rightarrow \sqrt{i/2} [1 - e^{i\eta} \sqrt{i/\pi\eta}] \quad \text{for } \eta \gg 1,$$

and

$$\mathcal{F}(\eta) \rightarrow \sqrt{2\eta/\pi} \quad \text{for } \eta \rightarrow 0.$$

For N infinite, the case of present interest, we obtain

$$\mathcal{R}_\pm = 2 \sum_{\nu} \frac{1 - \sqrt{2/i} \mathcal{F}(L_\nu)}{[1 + (D/b)e^{b\gamma_\nu}] \sqrt{2kbL_\nu}}. \quad (79)$$

The present expression for G , in terms of (79), may also be obtained by an asymptotic development (using the method of stationary phase) of the complete G of Section II in terms of J of (22a). In particular, for large values of L_ν we use the asymptotic form of \mathcal{F} in (79) and obtain

$$\mathcal{R}_\pm = \sqrt{\frac{2i}{\pi kb}} \sum \frac{e^{iL_\nu}}{[1 + (D/b)e^{b\gamma_\nu}]} \quad (79a)$$

which is proportional to J of (22a) with the previous t replaced by 1.

For the rare gas, \mathcal{R} reduces to the term for $\nu=0$:

$$\mathcal{R}_\pm \approx \sqrt{\frac{2}{1 \mp \sin \phi_0}} \frac{\{1 - \mathcal{F}(kb[1 \mp \sin \phi_0]) \sqrt{2/i}\}}{kb}. \quad (80)$$

Plots of the real and imaginary parts of this function for $kb=10$, and a fairly detailed discussion, are given in reference 1. In general for large kb [which is required for the heuristic form G of (76) to hold], we may neglect \mathcal{R}_- and approximate \mathcal{R}_+ by

$$\mathcal{R}_+ \approx \frac{2}{kb \cos \phi_0} = 2C_0. \quad (81)$$

The resultant form of (76) is then

$$G(\phi, \phi_0) = g(\phi, \phi_0) + g\left(\phi, \frac{\pi}{2}\right) G\left(\phi, \frac{\pi}{2}\right) \mathcal{R}_+ \\ = g(\phi, \phi_0) + \frac{g\left(\phi, \frac{\pi}{2}\right) g\left(\frac{\pi}{2}, 0\right) \mathcal{R}_+}{1 - g\left(\frac{\pi}{2}, \frac{\pi}{2}\right) \mathcal{R}_+}. \quad (82)$$

[See reference 1 (39).] The multiple scattering "correction" is significant only near grazing incidence ($\phi_0 \rightarrow \pi/2$), for which case

$$G(\phi, \phi_0) \rightarrow \frac{g(\phi, \phi_0)}{1 - g\left(\frac{\pi}{2}, \frac{\pi}{2}\right) 2C_0}; \quad (83)$$

this last result is also the limit obtained by the more general, and more rigorous Green's function treatment of the gas given in reference 3.

Another case of interest corresponds to the grating anomalies of Wood and Strong (which may also occur for the near periodic limit). Thus if

$$L_n = kb(1 - \sin \phi_0) - ib\gamma_n \rightarrow kb(1 - \sin \phi_0) - 2n\pi \rightarrow 0,$$

we need retain only the term $\nu=n$ of \mathcal{R}_+ of (79); ignoring the term in \mathcal{F} which remains finite in the limit, we obtain

$$\mathcal{R}_+ \approx \frac{2}{[1 + (D/b)e^{b\gamma_n}] \sqrt{2kbL_n}} \\ \rightarrow \frac{2}{kb \left[1 - \left(\sin \phi_0 + \frac{2n\pi}{kb}\right)^2\right]^{1/2}} = 2C_n, \quad (84)$$

essentially as in reference 1 (43). Using this result in (82), or in the analogous form of F_+ of (77), gives the "resonance form" discussed previously in detail.^{2,6} Thus since C_n is imaginary if an evanescent mode is near grazing ($\sin \phi_0 + 2n\pi/kb = 1 + |E| \rightarrow 1$), the first order terms of G may vanish for one value of ϕ , *i.e.*, when

$$\text{Im} \left\{ g(\phi, \phi_0) - 2C_n \left[g(\phi, \phi_0) g\left(\frac{\pi}{2}, \frac{\pi}{2}\right) \right. \right. \\ \left. \left. - g\left(\phi, \frac{\pi}{2}\right) g\left(\frac{\pi}{2}, 0\right) \right] \right\} = 0;$$

and G may show a marked maximum for another value of ϕ , i.e., when

$$\operatorname{Re} \left[1 - g \left(\frac{\pi}{2}, \frac{\pi}{2} \right) 2C_n \right] = 0.$$

Behavior of Reflection Coefficients Near Grazing Incidence

In reference 3 we showed that the reflection coefficients for rough surfaces consisting of a "gas distribution" of protuberances on a ground plane approached unity linearly as a function of the "grazing angle." We now apply (44) to show that the same results also hold for the general liquid distribution.

The function of primary interest is the "reflection amplitude" of $\langle U \rangle = [1 + 2C_0 F_{00}] \psi_0$ as involved in R of (65), or as is customary, the corresponding function for $\langle U \rangle$ divided by the incident field. Since the incident fields giving rise to $\langle U_+ \rangle$ and $\langle U_- \rangle$ are, respectively, $+\psi_0 = e^{ikr \cos(\phi - \pi + \phi_0)}$ and $-\psi_0$, we work with

$$\rho_{\pm} = \pm [1 + 2C_0 F(\phi_0, \pi - \phi_0)]. \quad (85)$$

Expanding F_{00} of (44) by a Neumann iteration procedure in terms of $p_{\nu\mu}$ gives

$$\begin{aligned} C_0 F_{00} &= C_0 p_{00} + C_0 \mathbf{S}_1 C_{\nu} p_{0\nu} p_{\nu 0} + C_0 \mathbf{S}_1 C_{\mu} p_{0\mu} \mathbf{S}_1 C_{\nu} p_{\mu\nu} p_{\nu 0} + \dots \\ &= \frac{C_0 f_{00}}{1 - C_0 f_{00}} + \mathcal{O} \left[\frac{C_0 f_{00} f_{\nu 0}}{(1 - C_0 f_{00})^2} \right], \end{aligned} \quad (86)$$

where the notation is to indicate the order of the dependence on the remaining terms containing functions of ϕ_0 .

We are concerned only with the range near grazing incidence, i.e.,

$$\phi = \frac{\pi}{2} - \epsilon \rightarrow \frac{\pi}{2}, \quad C_0 = \frac{1}{kB \cos \phi_0} \rightarrow \frac{1}{kB\epsilon} \rightarrow \infty. \quad (87)$$

In this range, we showed previously [see reference 3 (11)ff] that

$$\begin{aligned} f_- \left(\phi, \frac{\pi}{2} + \epsilon \right) &\rightarrow \mathcal{O}(\epsilon), \quad f_- \left(\left| \frac{\pi}{2} - \epsilon \right|, \pi - \phi \right) \rightarrow \mathcal{O}(\epsilon), \\ f_- \left(\left| \frac{\pi}{2} - \epsilon \right|, \frac{\pi}{2} + \epsilon \right) &\rightarrow \mathcal{O}(\epsilon^2); \end{aligned} \quad (88)$$

$$\begin{aligned} G &= \frac{F_- + F_+}{2} = \frac{2a_1 \cos \phi_0 \cos \phi}{1 - a_1(\mathcal{H}_0 + \mathcal{H}_2)} \\ &+ \frac{a_0 + 2a_1 \sin \phi_0 \sin \phi + a_0 a_1 [\mathcal{H}_2 + 2i\mathcal{H}_1(\sin \phi_0 + \sin \phi) - \mathcal{H}_0(1 + 2 \sin \phi_0 \sin \phi)]}{1 - a_0 \mathcal{H}_0 - a_1(\mathcal{H}_0 - \mathcal{H}_2) + a_0 a_1 (2\mathcal{H}_1^2 + \mathcal{H}_0^2 - \mathcal{H}_0 \mathcal{H}_2)} \end{aligned} \quad (94)$$

f_+ is non-vanishing. These general consequences of the symmetry of the problem hold for arbitrary scatterers.

Using (88) in (86), we find for horizontal polarization,

$C_0 f_{00} \rightarrow \mathcal{O}(\epsilon)$, and $C_0 f_{00} f_{\nu 0} \rightarrow \mathcal{O}(\epsilon)$. Consequently

$$C_0 F_{00} = \mathcal{O}(\epsilon), \quad \rho_- \rightarrow -1 + \mathcal{O}(\epsilon); \quad \phi_0 = \frac{\pi}{2} - \epsilon \rightarrow \frac{\pi}{2}. \quad (89)$$

On the other hand, for vertical polarization, we have

$$\begin{aligned} C_0 f_{00} \rightarrow \infty, \quad \frac{C_0 f_{00}}{1 - C_0 f_{00}} &= -\frac{1}{1 - \frac{1}{C_0 f_{00}}} = -1 + \mathcal{O}(\epsilon); \\ \frac{C_0}{(1 - C_0 f_{00})^2} &\rightarrow \frac{1}{C_0 f_{00}^2} = \mathcal{O}(\epsilon). \end{aligned}$$

Consequently

$$\begin{aligned} CF_{00} &\rightarrow -2 + \mathcal{O}(\epsilon), \quad \rho_+ \rightarrow -1 + \mathcal{O}(\epsilon), \\ \phi_0 &= \frac{\pi}{2} - \epsilon \rightarrow \frac{\pi}{2}. \end{aligned} \quad (90)$$

Thus, both reflection coefficients $R = |\rho|^2$ approach unity linearly as a function of the vanishing "horizon angle" ϵ , and both fields have undergone a phase change of π on reflection. [See reference 1 (44)ff and (96)ff and reference 3 (35) for further discussion.]

Circular Cylinders

As mentioned after (39), the forms (36) and (37) for the amplitudes in terms of the operator of (39), enable us to exploit the special procedures developed for the separable problems of circular and elliptic cylinders.^{6,7} In particular for circular cylinders, we use the Fourier series representations of the amplitudes

$$g(\phi, \phi_0) = \sum_{n=-\infty}^{\infty} a_n e^{in(\phi - \phi_0)}, \quad (91)$$

$$G(\phi, \phi_0) = \sum_{n=-\infty}^{\infty} A_n e^{in\phi}, \quad (92)$$

in (36), and obtain the infinite set of algebraic equations

$$\begin{aligned} A_n &= a_n [e^{-in\phi_0} + \sum A_m \mathcal{H}_{n-m}], \\ \mathcal{H}_{n-m} &= \mathbf{S} C_{\mu} [e^{-i(n-m)\phi_{\mu}} + e^{-i(n-m)(\pi - \phi_{\mu})}], \end{aligned} \quad (93)$$

which relate the "multiple scattered coefficients" A_n to their known⁶ single scattered values a_n .

The value of G obtained by retaining only the terms $n=0, \pm 1$ in (93) (i.e., all monopole and dipole contributions) was given originally in reference 1, (A1) of Rept. EM58.

which is identical in form with (95) of Rept. E28 in reference 6. Similarly, the various series expansions and more general closed forms of E28 apply equally for the

present more general distribution provided that the operator in \mathcal{H} is taken to be (39).

Comparison of (93) with analogous result obtained previously by separation of variables¹ leads to the identity.

$$\mathcal{H}_{n-m} = \sum_{s=1}^{\infty} \int H_{n-m}(k\eta) w_{0s}(\eta) [e^{ik\eta \sin \phi_0} (-1)^{n-m} + e^{-ik\eta \sin \phi_0}] d\eta \quad (95)$$

where $H_n = J_n + iN_n$ is the Hankel function of the first kind. More specifically, for even and odd order, we obtain

$$\begin{aligned} \mathcal{H}_{2n} &= 2 \mathbf{S} C_\mu \cos 2n\phi_\mu \\ &= 2 \sum_{s=1}^{\infty} \mathbf{S} H_{2n}(k\eta) w_{0s}(\eta) \cos(k\eta \sin \phi_0) d\eta \\ &= 2 \int H_{2n}(k\eta) w(\eta) \cos(k\eta \sin \phi_0) d\eta, \end{aligned} \quad (96)$$

$$\begin{aligned} \mathcal{H}_{2n+1} &= -2i \int C_\mu \sin(2n+1)\phi_\mu \\ &= -2i \sum \int H_{2n+1}(k\eta) w_{0s}(\eta) \sin(k\eta \sin \phi_0) d\eta \\ &= -2i \int H_{2n+1}(k\eta) w(\eta) \sin(k\eta \sin \phi_0) d\eta. \end{aligned} \quad (97)$$

The previous general considerations in regard to the behavior at grazing incidence, near-periodic resonances, etc., also apply to the present case. The primary range of interest for multiple scattering remaining to be considered is that of relatively close packed scatterers. For simplicity, we restrict consideration to the low-frequency range ($kB \ll 1$) and obtain closed forms for the "packing effects" for vertical polarization and perfect conductors.

Instead of obtaining the required limiting forms through a detailed treatment of the functions \mathcal{H} , we use the elementary procedure suggested in the analogous treatment of the periodic limit.⁸ Thus for $kB \rightarrow 0$, we approximate the Hankel functions by

$$H_{2n}(x) \rightarrow iN_{2n}(x) \rightarrow -\frac{i(n-1)!2^n}{\pi x^n}, \quad (98)$$

so that for normal incidence (96) reduces to

$$\mathcal{H}_{2n} \rightarrow -\frac{i2(2n-1)!2^{2n}}{\pi(kb)^{2n}} b \int_1^\infty \frac{w(b\eta)}{\eta^{2n}} d\eta, \quad (99)$$

where w is as in (12) with η replaced by $b\eta$.

In the near periodic limit ($b \rightarrow B$), we use w as in (12b) and obtain

$$\begin{aligned} \mathcal{H}_{2n} &\rightarrow -\frac{i2(2n-1)!2^{2n}}{\pi(kB)^{2n}} \sum_{s=1}^{\infty} \frac{1}{s^{2n}} \\ &= -\left(\frac{2\pi}{kB}\right)^{2n} \frac{B_n 2^{2n}}{2n\pi}, \end{aligned} \quad (100)$$

where the B_n 's are the Bernoullian numbers such that $B_1 = \frac{1}{6}$, $B_2 = \frac{1}{30}$, etc. This result is identical in form with that obtained previously^{6,8} for the periodic case ($B=b$).

On the other hand, near the gas limit, we use

$$\begin{aligned} b \int_1^\infty \frac{w(b\eta)}{\eta^{2n}} d\eta &\approx \frac{b}{B(2n-1)} - \sum_{\nu \neq 0} \frac{e^{b\gamma_\nu}}{(b + De^{b\gamma_\nu})\gamma_\nu} \\ &\equiv \frac{b}{B(2n-1)} - W, \quad |\gamma_\nu| \gg 1. \end{aligned} \quad (101)$$

Thus,

$$\begin{aligned} \mathcal{H}_{2n} &\rightarrow -\frac{i2(2n-1)!2^{2n}}{\pi(kb)^{2n}} \left[\frac{b}{B(2n-1)} - W \right] \\ &\approx H_{2n}(kb) \left[\frac{b}{B(2n-1)} - W \right]. \end{aligned} \quad (102)$$

In the following, we drop W for simplicity.

For perfect conductors and vertical polarization (*i.e.*, $\partial_n \Psi = 0$ at the surface), the single scattering coefficients for radius (a) small compared to wavelength, may be approximated by

$$a_0 = -\frac{i\pi(ka)^2}{4}; \quad a_n = \frac{i\pi n}{(n!)^2} \left(\frac{ka}{2}\right)^{2n}, \quad n = 0. \quad (103)$$

Using (106) and (102) in (93) (or in the more explicit forms of reference 6), we obtain the multiple scattered coefficients to lowest order in $k = 2\pi/\lambda \rightarrow 0$:

$$A_0 = a_0 = -\frac{i\pi(ka)^2}{4}, \quad (104)$$

$$A_1 = a_1 X = \frac{i\pi(ka)^2}{4} X, \quad (105)$$

where X depends only on a , b , and B (the radius, minimum separation, and average separation).

If we take into account only "dipole-dipole" interactions, then

$$X = \frac{1}{1 - a_1 \mathcal{H}_2}. \quad (106)$$

In the near periodic limit, this reduces to

$$X_p = \frac{1}{1 - \frac{\pi^2}{3} \left(\frac{a}{B}\right)^2}, \quad B \approx b, \quad (107)$$

and for the gas,

$$X_g = \frac{1}{1 - \frac{2b}{B} \left(\frac{a}{b}\right)^2} = \frac{1}{1 - \frac{2a^2}{Bb}}, \quad B \gg b. \quad (108)$$

⁸ W. V. Ignatowsky, *Arch. Math. Physik*, vol. 23, p. 193; 1913. Also, V. Twersky, "Elementary Function Representations of Schlömilch Series," Sylvania Electronic Defense Lab., Mountain View, Calif., Rept. EDL-E24; date, 1958. See equation (64)ff for extensions.

If we let $b = 2a$ in (108), so that the minimum separation or "exclusion region" (*i.e.*, the smallest region containing only one scatterer) equals the diameter of a cylinder, then X_g reduces to

$$X_g = \frac{1}{1 - \frac{a}{B}}. \quad (109)$$

so that the reflected intensity is greater for the gas than for the near periodic array when $a/B \ll 1$.

If we include dipole-dipole, dipole-octupole, and dipole-dipole-octupole interactions, then

$$X = \frac{(1 - a_3 \mathcal{C}_6)}{(1 - a_1 \mathcal{C}_2)(1 - a_3 \mathcal{C}_6) - a_1 a_3 \mathcal{C}_4^2}. \quad (113)$$

For the periodic case, this gives

$$X_p = \frac{1 - \frac{4}{7} \left(\frac{\pi^2}{3} \right)^3 \left(\frac{a}{B} \right)^6}{1 - \frac{\pi^2}{3} \left(\frac{a}{B} \right)^2 - \frac{4}{7} \left(\frac{\pi^2}{3} \right)^3 \left(\frac{a}{B} \right)^6 + \frac{79}{175} \left(\frac{\pi^2}{3} \right)^4 \left(\frac{a}{B} \right)^8}, \quad (114)$$

Thus, the near periodic and gas results are identical, if $a/B = 3/\pi^2$, and $X_p > X_g$ in the range $(a/B)_{\max} = \frac{1}{2} > a/B > 3/\pi^2$. However, we require $a/B \ll 1$ for the gas result to be valid, and in this range of $a/B < 3/\pi^2$, we have $X_g > X_p$.

In either case, X has the form $(1-v)^{-1}$, and the corresponding scattering amplitude can be written

$$\begin{aligned} G(\phi, 0) &= -\frac{i\pi k^2}{4} \left[a^2 - \frac{2a^2}{1-v} \cos \phi \right] \\ &= -\frac{i\pi k^2}{4} [\xi \eta - (\xi + \eta) \eta \cos \phi], \end{aligned} \quad (110)$$

which is of the form of the single scattered amplitude of an ellipse with axis

$$\xi = a \sqrt{\frac{1-v}{1+v}}, \quad \eta = a \sqrt{\frac{1+v}{1-v}}, \quad (111)$$

where ξ is perpendicular to the distribution. The corresponding reflection amplitude has the form

$$\rho = -\frac{2}{kB} G(\pi, 0) = -\frac{i\pi^2 a^2}{\lambda B} (1 + 2X), \quad (112)$$

and for the gas,

$$X_g = \frac{1 - \frac{4b}{B} \left(\frac{a}{b} \right)^6}{1 - \frac{2b}{B} \left(\frac{a}{b} \right)^2 - \frac{4b}{B} \left(\frac{a}{b} \right)^6 + \frac{20}{3} \left(\frac{b}{B} \right)^2 \left(\frac{a}{b} \right)^8}. \quad (115)$$

If $b = 2a$, the gas result reduces to

$$X_g = \frac{1 - \frac{1}{8} \left(\frac{a}{B} \right)}{1 - \frac{9}{8} \left(\frac{a}{B} \right) + \frac{5}{92} \left(\frac{a}{B} \right)^2}; \quad (116)$$

thus, the a/B term in the denominator has been increased [relative to (109)] by the octupole contribution.

The above comparison between the two limiting forms of the general distribution indicates that the effects of "packing" are relatively different for the periodic and gas cases for $a/B \ll 1$ —even in the low frequency limit for which it has been customary to interchange the results. However, the present treatment is merely illustrative; for detailed considerations one should use more than the leading term of (101).

Modified WKB Methods for the Propagation and Scattering of Electromagnetic Waves*

D. S. SAXON†

Summary—A new formulation of scattering and propagation problems has been developed using Green's functions with essentially correct local behavior. This formulation, which is exact, yields the familiar WKB result as its zeroth approximation. Higher order corrections depend on the spatial variation of the local index of refraction. The convergence is rapid if these variations are gradual. By way of illustration, the method has been applied to the propagation of electromagnetic waves in an isotropic stratified medium and to the scattering of waves by an inhomogeneity. In the former case, the problem of total reflection, as by the ionosphere, has been studied and corrections to the usual WKB phase shift have been obtained. These corrections arise as a consequence of deviations of the effective dielectric constant from linearity in the neighborhood of the turning point, and also because of the slow approach of the height-gain function to its asymptotic value far from the turning point. In the latter example, approximate expressions are obtained for both the scattering amplitude and the close-in electric field which are valid when the index of refraction of the scattering center is close to unity. Polarization effects are examined explicitly.

I. DESCRIPTION OF THE METHOD

PROBLEMS involving the propagation of electromagnetic waves in inhomogeneous media, or the scattering of electromagnetic waves by an inhomogeneity, are so difficult that exact solutions are unknown except in the simplest and most highly idealized cases. Even in such cases, the solutions are usually too complicated to be generally useful; for example, the exact expressions for the scattering of light by a dielectric sphere are tractable only for spheres not too large compared to a wavelength. The entire history of these problems, from Huygens, Fresnel and Kirchhoff to the present, is accordingly dominated by attempts to develop suitable approximation methods. In the following we present still one more such attempt.

In a certain sense our procedure is a generalization of the method of Fermat, by which we mean of the geometrical optical approximation or, in wave theory, of the WKB approximation. The essential idea is the following: the wave equation is rewritten as an integral equation by introducing a Green's function which has essentially correct local behavior, that is, which describes propagation along a geometrical optical path with correct local wave number.¹ This integral equation, which is exact, yields the WKB result as its zeroth approximation. Higher order corrections depend on the variation of the local wave number along the ray and on the curvature of the ray. This procedure, while physically appealing, is very complicated mathematically.

However, choice of a less complex Green's function is always possible. In particular, if a simple straight line approximation to the geometrical optical rays is introduced, one obtains a correspondingly simpler, but still exact, integral equation which has proved useful in the formulation of quantum mechanical scattering problems.^{2,3}

The actual construction of such integral equations is carried out by first writing down a Green's function with the desired character and then finding the differential equation it satisfies, a straightforward matter. To the extent that the Green's function is well chosen, its differential equation differs little from that satisfied by the actual fields. In any case, the integral equation follows directly upon application of Green's theorem. Of course, in the electromagnetic applications we have in mind, there are often additional complications associated with the vector nature of the field, but as it turns out these can be handled in a straightforward way. It should be mentioned that whenever a physical problem reduces to the solution of an ordinary differential equation with a turning point, our method reduces to that of Langer.⁴

Admittedly, these descriptive remarks are somewhat sketchy. We have chosen to fill out the picture by describing the application of the method to two illustrative examples; namely, to the problem of total reflection in an isotropic stratified medium, and to the scattering of waves by an isolated inhomogeneity.

II. TOTAL REFLECTION IN AN ISOTROPIC STRATIFIED MEDIUM⁵

We consider first the total reflection of electromagnetic waves in an isotropic stratified medium. By an isotropic stratified medium we mean one characterized by constitutive parameters depending only on a single rectangular coordinate, the z -coordinate, for example. As is well-known,⁶ the solution of the vector problem can be expressed in terms of a pair of scalar functions, one for the case in which the electric field is polarized perpendicular to the z -axis, the other where it is the

² D. S. Saxon and L. I. Schiff, *Nuovo Cim.*, vol. 6, p. 614; 1957.

³ W. M. Brown, Ph.D. dissertation, Dept. of Physics, Univ. of California (to be published).

⁴ R. E. Langer, *Trans. Amer. Math. Soc.*, vol. 37, p. 397, 1935; vol. 67, p. 461, 1949.

⁵ D. S. Saxon, "Total Reflection of Electromagnetic Waves in a Stratified Isotropic Medium," System Corp. of America, Tech. Rept. 3-15-59; unpublished.

⁶ H. Bremmer, "Encyclopedia of Physics," Springer-Verlag, Berlin, Ger., vol. 16; 1958. This is an excellent summary which gives references to earlier work.

* This work was sponsored in part by Republic Aviation Corp., Farmingdale, L. I., N. Y., and the U. S. National Science Foundation.

† Dept. of Physics, University of California, Los Angeles, Calif.

¹ D. S. Saxon, *Phys. Rev.*, vol. 107, p. 871; 1957.

magnetic field which is so polarized. In each case this scalar function, which is proportional to the height-gain function and which we denote by $\psi(z)$, satisfies a differential equation of the form

$$\frac{d^2\psi}{dz^2} + \kappa^2(z)\psi = 0, \quad (1)$$

where $\kappa(z)$ is the local wave number for propagation in the z direction. It depends on the effective index of refraction and on the obliquity of the rays. For the case of E -polarization, at frequency ω , for example,

$$\kappa^2(z) = \omega^2 \mu_0 \tilde{\epsilon}(z) - \gamma^2,$$

where $\tilde{\epsilon}(z)$ is the effective inductive capacity, and γ the transverse wave number.

We restrict our attention to the case in which $\kappa^2(z)$ is real, attains the asymptotic value k^2 as z recedes to minus infinity and becomes negative when z exceeds z_0 , which we shall call the turning point. Further, in the immediate neighborhood of the turning point, we shall assume that $\kappa^2(z)$ is linear in z , that is, that

$$\kappa^2(z) \sim (z_0 - z); \quad z \simeq z_0. \quad (2)$$

Under these conditions a wave incident from $z = -\infty$ is totally reflected since real propagation is impossible past the turning point. The behavior of a geometrical optical ray is roughly indicated in Fig. 1 for a given initial direction (γ , and therefore z_0 , depend on this direction).

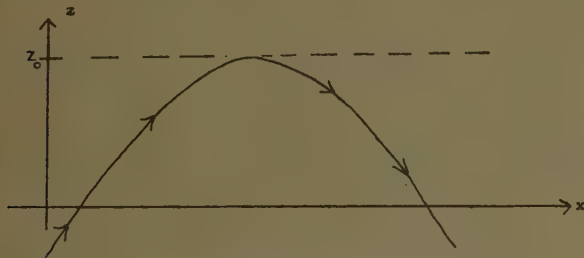


Fig. 1—Geometrical optical path for a totally reflected ray.

To apply our method, we must now construct an appropriate Green's function, that is, one which has correct local behavior. In particular, in this example, we mean correct behavior both at the turning point and asymptotically. We assert that the following is such a Green's function:⁴

$$G(z, z') = G(z', z) = \frac{\pi}{8} C_1(z_>) C_2(z_<), \quad (3)$$

where $z_>$ means z or z' , whichever is greater, while $z_<$ means the smaller of the two and where

$$\begin{aligned} C_1(z) &= e^{2\pi i/3} Z^{1/2} \kappa^{-1/2} H_{\frac{1}{2}}^{(1)}(Z), \\ C_2(z) &= e^{\pi i/6} Z^{1/2} \kappa^{-1/2} \{ H_{\frac{1}{2}}^{(1)}(Z) + 2e^{-\pi i/3} H_{\frac{1}{2}}^{(2)}(Z) \}, \end{aligned} \quad (4)$$

with

$$Z(z) = \int_{z_0}^z \kappa(z) dz. \quad (5)$$

Recalling that $\kappa^2(z)$ is linear near the turning point, as indicated by (2), the phase of Z is chosen such that

$$\begin{aligned} Z &= |Z| e^{i\pi/2}, & z > z_0, \\ Z &= |Z| e^{-i\pi}, & z < z_0. \end{aligned} \quad (6)$$

That the Green's function of (3) has the desired behavior can be seen by examining the characteristics of the functions C_1 and C_2 . First we look at their behavior in the neighborhood of the turning point. Here it follows that

$$Z \sim (z_0 - z)^{3/2} \sim \kappa^3,$$

and hence, C_1 and C_2 are exact solutions of the strictly linear turning point problem. Next we look at their asymptotic behavior. For $z \rightarrow -\infty$, using (6), and the known continuation properties of the Bessel functions, we obtain

$$\begin{aligned} C_1(z) &\simeq 2\sqrt{\frac{2}{\pi\kappa}} \cos(|Z| - \pi/4), \\ C_2(z) &\simeq -2\sqrt{\frac{2}{\pi\kappa}} \sin(|Z| - \pi/4), \end{aligned} \quad (7)$$

while for $z \rightarrow \infty$,

$$\begin{aligned} C_1(z) &\simeq \sqrt{\frac{z}{\pi|\kappa|}} e^{-|Z|}, \\ C_2(z) &\simeq 2\sqrt{\frac{z}{\pi|\kappa|}} e^{|Z|}, \end{aligned} \quad (8)$$

and hence, C_1 and C_2 are also seen to yield the correct asymptotic solutions.

The differential equation satisfied by the Green's function of (3) is readily shown to be

$$\frac{d^2G}{dz^2} + (\kappa^2 - \lambda)G = -\delta(z - z'), \quad (9)$$

where

$$\lambda(z) = z^{-1/6} \kappa^{1/2} \frac{d^2}{dz^2} (z^{1/6} \kappa^{-1/2}). \quad (10)$$

Consequently, applying Green's theorem to G and the height-gain function ψ , we obtain after some manipulation,

$$\psi(z) = C_1(z) + \int_{-\infty}^{\infty} \lambda(z') \psi(z') G(z, z') dz', \quad (11)$$

which is the fundamental equation of our treatment.

In words, we have exhibited the height-gain function as the familiar WKB solution,⁶ $C_1(z)$, plus a correction integral, the magnitude of which depends on the extent to which $\lambda(z)$ departs from zero. It follows from the definitions that over any region (which may include the turning point) for which $\kappa^2(z)$ is strictly linear, λ

is identically zero, while for $z \rightarrow -\infty$, λ vanishes as $1/z^2$.⁷ Thus the correction to the WKB result is seen to contain two contributions:

- 1) The contribution arising from deviations from linearity near the turning point;
- 2) The contribution arising from the gradual approach of the solution to its asymptotic form far from the turning point.

In the usual treatment it is assumed that κ^2 remains linear out to large enough distances from the turning point that the asymptotic solutions become valid. If so, λ is small everywhere and the correction is presumably negligible. The present exact formulation thus furnishes a precise basis for these assumptions. More generally, with the aid of iterative, or other appropriate techniques, it makes it possible to calculate corrections to the WKB result.

We now look at the asymptotic form in order to make explicit the effect of such corrections. From (3) and (11), we have for $z \rightarrow -\infty$,

$$\psi(z) \simeq C_1(z) + C_2(z) \cdot \frac{\pi}{8} \int_{-\infty}^{\infty} \lambda(z) \psi(z) C_1(z) dz$$

where, introducing the asymptotic forms of C_1 and C_2 , it follows that ψ can be written in the form,

$$\psi(z) \simeq 2 \sqrt{\frac{2}{\pi \kappa}} \frac{\cos(Z - \delta + \pi/4)}{\cos \delta}, \quad (12)$$

where

$$\delta = \tan^{-1} \frac{\pi}{8} \int_{-\infty}^{\infty} \lambda(z) \psi(z) C_1(z) dz. \quad (13)$$

Physically, the phase shift δ is seen to represent a change in the apparent location of the reflecting layer.

The result of (13) is exact, of course, but it involves the unknown function ψ . However, if λ is not too large, we expect that an iterative solution of the fundamental integral equation will converge rapidly. Under this condition, the phase shift is then given to first order by

$$\delta \sim \tan^{-1} \frac{\pi}{8} \int \lambda(z) C_1^2(z) dz. \quad (14)$$

If $\kappa^2(z)$ is linear over a sufficiently extended region, as assumed in the usual WKB method, this result can be simplified to⁵

$$\delta \simeq \tan^{-1} \int_{-\infty}^{z_0} \frac{\lambda}{2\kappa} dz. \quad (15)$$

As might be expected, the latter can also be derived more directly by a suitable transformation of the differential equation satisfied by ψ .

⁷ Since the integrand is exponentially small for large positive z , the behavior of λ there is a matter of indifference.

We have thus obtained explicit expressions for the correction to the WKB result. We shall not elaborate on the application of this formulation here. We only remark that its utility in correcting the WKB approximation has been discussed in Saxon,⁵ where a detailed treatment of this problem is given by consideration of some special examples, simple enough that exact solutions can be obtained.

III. SCATTERING BY AN ISOLATED INHOMOGENEITY³

As our second example, we consider the scattering of electromagnetic waves by an isolated inhomogeneity in empty space. For simplicity we take μ to equal one everywhere (Gaussian units), so that the inhomogeneity extends over the finite domain in which ϵ differs from unity. Assuming harmonic time dependence and eliminating \mathbf{H} from Maxwell's equations, we obtain as the equation to be solved,

$$(\nabla^2 + k^2)\mathbf{E} = -(\epsilon - 1)k^2\mathbf{E} + \nabla(\nabla \cdot \mathbf{E}), \quad (16)$$

where k is the free-space wave number. As in reference (1), we introduce a scalar Green's function

$$F(\mathbf{r}, \mathbf{r}') = F(\mathbf{r}', \mathbf{r}) = -\frac{e^{iS(\mathbf{r}, \mathbf{r}')}}{4\pi |\mathbf{r} - \mathbf{r}'|}, \quad (17)$$

where the phase factor S will later be chosen to give some approximation to the local propagation characteristics. For now we assume only that $S(\mathbf{r}, \mathbf{r}) = 0$, or more precisely, that

$$\lim_{\mathbf{r}' \rightarrow \mathbf{r}} \frac{S(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} = C(\mathbf{r}).$$

We further assume that as $\mathbf{r} = n\mathbf{r} \rightarrow \infty$, with \mathbf{r}' finite, the propagation proceeds along \mathbf{n} with the free space propagation constant, *i.e.*, that

$$\mathbf{r} = n\mathbf{r} \rightarrow \infty, \quad \nabla S = k\mathbf{n} + 0(1/r). \quad (18)$$

It is easily verified that F satisfies the differential equation

$$\nabla^2 F + (\nabla S)^2 F = \delta(\mathbf{r} - \mathbf{r}') + iF\rho^2 \nabla \cdot \left(\frac{\nabla S}{\rho^2} \right), \quad (19)$$

where

$$\rho = |\mathbf{r} - \mathbf{r}'|. \quad (20)$$

Applying Green's theorem to F and each component of \mathbf{E} , it can be shown that \mathbf{E} satisfies the exact integral equation

$$\begin{aligned} \mathbf{E} + E_0 e^{ikn_0 \cdot \mathbf{r} + i\delta_0(\mathbf{r})} \\ + \int \left[(\nabla' S)^2 - \kappa^2(\mathbf{r}') - i\rho^2 \nabla' \cdot \frac{\nabla' S}{\rho^2} \right] \mathbf{E}(\mathbf{r}') F(\mathbf{r}, \mathbf{r}') d^3\mathbf{r}' \\ + \int [1 - \epsilon(\mathbf{r}')] \mathbf{E}(\mathbf{r}') \cdot \nabla' \nabla' F(\mathbf{r}, \mathbf{r}') d^3\mathbf{r}', \end{aligned} \quad (21)$$

where we have introduced the local wave number $\kappa(\mathbf{r}) = k\sqrt{\epsilon}$, and where

$$\delta_0(\mathbf{r}) = -k\mathbf{n}_0 \cdot \mathbf{r} + \lim_{r' \rightarrow \infty} [S(\mathbf{r}, -\mathbf{n}_0\mathbf{r}') - k\mathbf{r}']. \quad (22)$$

Thus δ_0 represents the change in phase of a ray incident from infinity along \mathbf{n}_0 and arriving at \mathbf{r} compared to its value for free-space propagation.

Were the last integral of (21) not present, the components of \mathbf{E} would be uncoupled and each would satisfy the scalar integral equation treated in the literature.^{1,2} The last term, the *mixing integral*, is the entire source of the difference between vector and scalar problems. In the first place, it supplies precisely the terms required to annihilate the longitudinal components of the scattered radiation field and in the second, it gives rise to polarization effects.

We now remark on the choice of S . The simplest choice is $k|\mathbf{r} - \mathbf{r}'|$ corresponding to free-space propagation with corrections arising as a consequence of the deviation of ϵ from unity. Only if these deviations are small, and do not extend over too large a region, will these corrections be small. In this way one is led to the Kirchhoff-Born approximation. A much better choice is $(\nabla S)^2 = \kappa^2(\mathbf{r})$, corresponding to propagation with the correct local wave number along the geometrical optical ray paths, that is, to the WKB approximation, with corrections arising from changes in wave number along the ray and from the curvature of the ray. These corrections will be small if ϵ does not change too abruptly. This description is mathematically very complicated, however, and for simplicity we actually choose straight line ray paths with WKB phase instead of optical ray paths. That is, we write

$$S(\mathbf{r}, \mathbf{r}') = \int_{\mathbf{r}'}^{\mathbf{r}} \kappa(\mathbf{r}) d\mathbf{s} = \int_0^p \kappa(\mathbf{r}' + \mathbf{e}s) d\mathbf{s}, \quad (23)$$

where \mathbf{e} is a unit vector along the straight line from \mathbf{r}' to \mathbf{r} . With this choice

$$\delta_0 = \int_0^\infty [\kappa(\mathbf{r} - \mathbf{n}_0 s) - k] d\mathbf{s}.$$

The integral (21) is, of course, still exact.

If ϵ does not deviate too much from unity, we expect that an iteration treatment based on the integral equation will be rapidly convergent and that the first term will yield a good approximation to the scattering amplitude. To this order, the subsequent analysis (which is complicated) is almost exactly that of reference 2, the mixing integral contributes nothing but the terms required for transversality. Thus, for example, for large angle scattering ($kR\theta^2 \gg 1$), the amplitude A for waves scattered in the direction \mathbf{n} is given by³

$$A = -\mathbf{n} \times \mathbf{n} \times \frac{E_0}{4\pi} \int (\kappa^2 - k^2) e^{i(\mathbf{q} \cdot \mathbf{r} + \delta_0 + \delta_+)} d^3\mathbf{r}, \quad (24)$$

where $\mathbf{q} = k(\mathbf{n}_0 - \mathbf{n})$ and where δ_+ , the phase shift from a point \mathbf{r} along the straight line path to infinity in the direction of observation \mathbf{n} , is given by

$$\delta_+ = \int_0^\infty [\kappa(\mathbf{r} + \mathbf{n}s) - k] d\mathbf{s}.$$

Techniques for approximately evaluating integrals, such as that of (24), have been developed by Brown.³

Note that there are no polarization corrections to this order. To estimate them consistently, an improved wave function must be used inside the scatterer, the term of interest being generated by the mixing integral. After an integration by parts this becomes

$$\int F \nabla' \nabla' \cdot [(1 - \epsilon) \mathbf{E}(\mathbf{r}')] d^3\mathbf{r}',$$

from which estimates of each component of the secondary field can be obtained. If ϵ is slowly varying, it is easy to see that the dominant contribution is polarized in the direction of incidence.

As a specific example, the polarization corrections have been roughly estimated by Brown³ for scattering from a dielectric sphere of radius R . In particular, for scattering in the direction of polarization of the incident wave (where the first order result vanishes identically), this scattered field, polarized in the direction of incidence, is given by

$$A = \frac{(\epsilon - 1)^2 R}{8} \left\{ \left[\frac{ikR(\epsilon - 1)}{\sqrt{2}} + 1 \right] e^{ikR(3-\epsilon)/\sqrt{2}} + e^{-ikR\sqrt{2}} \right\}. \quad (25)$$

Note that this secondary scattered field is of the order $(\epsilon - 1)^2$, while the primary scattered field of (24) is of the order $(\epsilon - 1)$. It is also of interest to remark that the first term of (25) is a volume term, while the second and third terms are surface terms.

IV. CONCLUDING REMARKS

In the above we have described an attempt to develop a new formulation of electromagnetic propagation and scattering problems. The method has been illustrated by considering its application to the problems of total reflection in an isotropic stratified medium and of scattering by an isolated inhomogeneity. Application of the method to the study of propagation in a nonisotropic stratified medium for both the transmission and total reflection cases is now under consideration. In addition, the possibility of reformulating the scattering problem so that the polarization effects become explicit is also being investigated.

Interaction of Electromagnetic Waves with Some Natural Surfaces*

WILLIAM H. PEAKE†

Summary—The problem of the interaction of electromagnetic radiation with nonuniform surfaces (terrain, roadways, etc.) is of interest for predicting the apparent temperature of radiometers or radio telescopes. In this paper, the interaction is described by the differential scattering coefficients of the surface, in terms of which one may express such parameters of the surface as the radar cross section, the absorption coefficient, the albedo, etc. By making use of the reciprocity properties of the differential scattering coefficients, Kirchhoff's radiation law is derived in its most general form, which takes account of both the angular dependence and the polarization properties of the emitted radiation. Thus, the emissivity of the surface can also be expressed in terms of the scattering coefficients. General formulas for apparent surface temperature are obtained and are used to calculate the apparent temperature of an asphalt roadway and a vegetation-covered surface. The predicted temperatures are found to be in reasonably good agreement with measurements of Britt, Tolbert and Straiton¹ at 4.3 mm wavelength.

INTRODUCTION—THE DIFFERENTIAL SCATTERING COEFFICIENTS

IN RECENT years the problem of predicting the apparent temperature of terrain surfaces at microwave frequencies has become of interest in calculating the performance of radiometers and radio telescopes. It is the purpose of the paper to show how the scattering properties of the surface determine its apparent temperature. The actual calculation of these scattering properties for the general nonuniform surface is an extremely difficult one and is not considered here. References to the subject may be found in two recent papers by Twersky.^{2,3}

The interaction of electromagnetic waves with natural surfaces may most conveniently be described in terms of the differential scattering coefficients, for which several slightly different definitions exist. Here a generalization of one used in terrain return work is adopted.⁴ Consider an infinite surface forming, on the average, a plane, although over small regions it may have any complexity of structure (see Fig. 1). If radiation of intensity I_o (watts per meter²) falls at an angle of incidence θ_o and azimuth ϕ_o on a given element of surface area S , and the intensity of the scattered radiation in the direction θ_s, ϕ_s at a distance R from S is I_s , then the differential scatter-

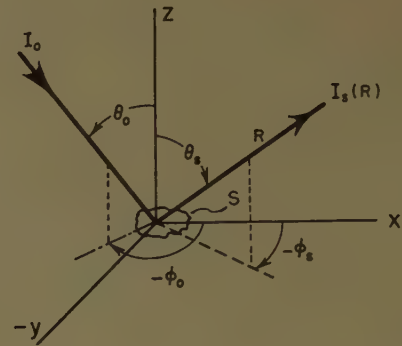


Fig. 1—Geometry of the scattering problem.

ing coefficient $\gamma(\theta_o, \phi_o; \theta_s, \phi_s)$ is defined by

$$\gamma(\theta_o, \phi_o; \theta_s, \phi_s) = (4\pi R^2 I_s) / (I_o S \cos \theta_o). \quad (1)$$

For a general surface, the scattered radiation I_s and thus γ itself consists of two parts, a "specular" or coherent part, and a "diffuse" or incoherent part.^{2,3} The specular part gives a γ which has the character of a δ function. (E.g., for a perfectly flat surface $\gamma(\theta_o, \phi_o; \theta_s, \phi_s) = 4\pi |R_i|^2 \csc \theta_{sp} \delta(\theta_s - \theta_{sp}) \delta(\phi_s - \phi_{sp})$ where θ_{sp}, ϕ_{sp} represent the specular direction and $|R_i|^2$ is the reflection coefficient.) For the diffuse part, I_s is proportional to S and inversely proportional to R^2 at large distances so that γ is independent of R and S . It is clear that S must be much larger than any significant structural feature of the surface if γ is also to be independent of the particular area at which the illumination is directed.

For electromagnetic radiation, the polarization of both incident and scattered radiation must be specified. Although any two orthogonal polarization states are suitable, the customary⁵ vertical (v) and horizontal (h) polarization states are used here. Thus, in what is to follow, the scattering coefficient will be written in the form $\gamma_{ij}(o, s)$; here the first subscript indicates the polarization state of the incident radiation and the second subscript indicates which polarization state of the scattered radiation is being considered; the first letter or pair of angles in the brackets indicates the direction from which the radiation is incident and the second letter or pair of angles indicates the direction into which the radiation is scattered; the letter (o) is an abbreviation for the angles θ_o, ϕ_o and the letter (s) an abbreviation for the angles θ_s, ϕ_s . It is a consequence of the reciprocity theorem that the scattering coefficients satisfy the four relations,

$$\cos \theta_o \gamma_{ij}(o, s) = \cos \theta_s \gamma_{ji}(s, o). \quad i, j \text{ either } v \text{ or } h. \quad (2)$$

* The research reported in this paper was sponsored in part by a contract between The Ohio State University Res. Found. and Wright Air Dev. Cen., under Contract AF 33 (616)-6158.

† Antenna Lab., Dept. of Electrical Engrg., The Ohio State University, Columbus, Ohio.

¹ A. W. Straiton, C. W. Tolbert, and C. O. Britt, "Apparent temperature distributions of some terrestrial materials and the sun at 4.3 mm wavelength," *J. Appl. Phys.*, vol. 29, pp. 776-782; May, 1958.

² V. Twersky, "On scattering and reflection of electromagnetic waves by rough surfaces," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-5, pp. 81-90; January, 1957.

³ V. Twersky, "On scattering and reflection of sound by rough surfaces," *J. Acoust. Soc. Am.*, vol. 29, pp. 209-225; February, 1957.

⁴ R. C. Taylor, "Terrain return measurements at X , K_u and K_a band," 1959 IRE NATIONAL CONVENTION RECORD, pt. 1, pp. 19-26.

⁵ D. E. Kerr, ed., "Propagation of Short Radio Waves," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 396-397; 1951.

To conclude this section two parameters common in radiometry will be defined in terms of the γ 's. The albedo of a surface usually refers to the fraction of incident radiation of random polarization rescattered by the surface for a band of frequencies in the optical region. Here the albedo A is defined as the fraction of the power incident on the surface from the direction $\theta_o\phi_o$ (at a specific polarization and frequency) that is rescattered. With this definition the albedo is

$$A(\theta_o, \phi_o) = \int \frac{I_s R^2 d\Omega_s}{SI_o \cos \theta_o}.$$

Here, and elsewhere unless otherwise specified, integration is over the upper hemisphere. Thus, from (1), taking account of the polarization properties of the surface, the albedo $A_i(\theta_o\phi_o)$ for incident radiation of specific polarization is

$$A_i(\theta_o\phi_o) = (4\pi)^{-1} \int [\gamma_{ii}(o, s) + \gamma_{ij}(o, s)] d\Omega_s, \quad (3)$$

$$i = h \text{ or } v$$

$$j = v \text{ or } h.$$

Another surface parameter of interest is the absorption coefficient (a), defined here as the fraction of power (of a given polarization and frequency) incident on the surface from the direction $\theta_o\phi_o$ that is absorbed by the surface. Clearly, for the types of surfaces considered here, which are assumed to be thick enough so that no energy is transmitted through the body they define, the absorption coefficient is just unity minus the albedo,

$$a_i(\theta_o\phi_o) = 1 - A_i(\theta_o\phi_o) \quad i = h \text{ or } v. \quad (4)$$

KIRCHHOFF'S LAW

In addition to the parameters discussed in the Introduction, a complete description of the interaction of electromagnetic waves with nonuniform surfaces requires a specification of their emissivities so that the radiometric properties (apparent surface temperatures) may be calculated. There do exist certain special surfaces for which the emissivity may be calculated directly (for example,⁶ a perfectly flat surface of arbitrary dielectric constant). However, for nonuniform surfaces in general, it appears necessary to invoke Kirchhoff's radiation law in order to calculate the emissivity from the scattering coefficients γ . An understanding of this relation between the scattering coefficients and the apparent temperature of natural surfaces is particularly important at microwave frequencies. Here the scattering coefficients can be measured rather easily, whereas it may be difficult, if not impossible, to separate the several contributions to the apparent temperature by measurement of thermal radiation alone.

Unfortunately, the usual statement of Kirchhoff's Law is not adequate for the purpose. To remedy this situation, von Fragstein⁷ has recently given a critical discussion of the form which the law must take for both smooth and diffuse surfaces, and the relation between it and the reciprocity properties of the scattering coefficients. However, his discussion was for the optical case and ignored the polarization properties of the surface. It will thus be necessary to extend his results to take account of these polarization effects in order to provide a statement of Kirchhoff's Law in its most general form.

Before carrying out the derivation, it will be convenient to define the parameters that will be used to characterize the ability of a surface to emit thermal radiation.

These are, in the terminology of the review article by Rutgers,⁸ the emission coefficients $e_h(\theta_o\phi_o)$, where

$$e_h(\theta_o\phi_o) = \frac{\text{Power emitted with horizontal polarization by unit area of surface into element of solid angle } d\Omega_o \text{ in the direction } \theta_o\phi_o.}{\text{Power emitted with horizontal polarization by unit area of black body at same temperature into same element of solid angle in same direction.}} \quad (5)$$

An analogous definition holds for $e_v(\theta_o\phi_o)$, for vertically polarized radiation. For a black body,

$$e_h(\theta_o\phi_o) = e_v(\theta_o\phi_o) = 1.$$

It is to be understood that throughout this report the thermal radiation referred to occupies a narrow band of frequencies Δf .

To obtain the desired relation between the emission coefficient and the absorption coefficient, defined earlier in (4), consider a surface (see Fig. 2) in temperature equilibrium with black body radiation in the half-space above it. Under this equilibrium condition, it will be assumed that just as much energy of a given polarization leaves the surface in a given direction as falls upon it from the same direction with the same polarization.

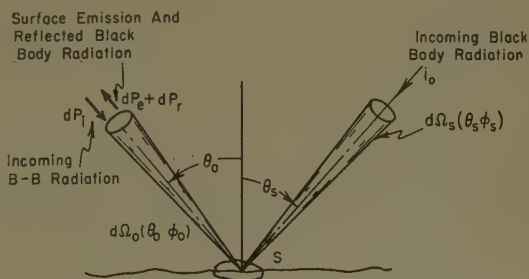


Fig. 2—Geometry for derivation of Kirchhoff's law.

⁷ C. von Fragstein, "Über die Formulierung des Kirchhoffschen Gesetzen, und ihre Bedeutung für eine Zweckmassige Definition von Remissionzahlen," *Optik*, vol. 12, pp. 60-70; 1955.

⁸ G. A. W. Rutgers, "Temperature radiation of solids," in "Handbuch der Physik," Springer-Verlag, Berlin, Ger., vol. XXVI, p. 129; 1958.

⁶ Max Planck, "Theory of Heat Radiation," Blakiston, Philadelphia, Pa.; 1914. (Transl. M. Masius.)

To state this assumption quantitatively, let i_o be the power density per unit solid angle (watts-meter⁻²-steradian⁻¹) of the black body radiation in a specific polarization state. (Thus, the total power density per unit solid angle of the black body radiation is $2i_o$.) Then i_o is a function of frequency and the equilibrium temperature but not of the direction or polarization of the radiation. It refers to the energy in the frequency interval Δf . The power dP_i incident with horizontal polarization on an element of surface of area S from a range of solid angle $d\Omega_o$ in direction $\theta_o\phi_o$ is

$$dP_i = i_o d\Omega_o S \cos \theta_o.$$

The power leaving the surface is the sum of two parts, the thermal emission from the surface, and the external black body radiation reflected by the surface. The power dP_e emitted by the same element of surface into the same range of solid angle with horizontal polarization is, from the definition of emission coefficient,

$$dP_e = e_h(\theta_o\phi_o) i_o d\Omega_o S \cos \theta_o.$$

The power reflected by the surface into $d\Omega_o$ in the direction $\theta_o\phi_o$ with horizontal polarization, caused by black body radiation coming from the range of solid angle $d\Omega_s$ in the direction $\theta_s\phi_s$ is, from (1),

$$d^2P_r = i_o d\Omega_o S \cos \theta_s [\gamma_{vh}(s, o) + \gamma_{hh}(s, o)] \frac{d\Omega_o}{4\pi}.$$

Thus the power dP_r reflected into $d\Omega_o$ with horizontal polarization is

$$dP_r = \frac{i_o d\Omega_o S}{4\pi} \int [\gamma_{hh}(s, o) + \gamma_{vh}(s, o)] \cos \theta_s d\Omega_s.$$

Assuming that the incident power of given polarization is balanced by the sum of the emitted and reflected power of the same polarization,

$$dP_i = dP_e + dP_r.$$

Consequently,

$$1 = e_h(\theta_o\phi_o) + \frac{1}{4\pi} \int \frac{\cos \theta_s}{\cos \theta_o} [\gamma_{hh}(s, o) + \gamma_{vh}(s, o)] d\Omega_s.$$

By the reciprocity relations of (2) this becomes

$$1 = e_h(\theta_o\phi_o) + \frac{1}{4\pi} \int [\gamma_{hh}(o, s) + \gamma_{hv}(o, s)] d\Omega_s.$$

The third term in this equation is just the albedo defined by (3), which is related to the absorption coefficient by (4). Thus, substituting, one finds

$$e_h(\theta_o\phi_o) = a_h(\theta_o\phi_o). \quad (6a)$$

A similar relation,

$$e_v(\theta_o\phi_o) = a_v(\theta_o\phi_o), \quad (6b)$$

is easily derived for vertical polarization.

Eqs. (6a) and (6b) constitute the desired generalization of Kirchhoff's law, taking account of both the polarization properties of the surface and the angular dependence of the coefficients. Furthermore through (3) and (4), the emission coefficients can be found from the scattering coefficients γ of the surface.

The correct standard form of Kirchhoff's law, which usually refers to the somewhat imprecise notion of the "absorptivity" or "absorbing power" of the surface for radiation entering from the entire upper hemisphere, may be obtained by multiplying both sides of (6a) or (6b) by $\cos \theta_o$ integrating over the hemisphere and averaging over the two polarizations.

APPARENT SURFACE TEMPERATURES

In practice, real terrain surfaces are not in equilibrium with black-body radiation in the half-space above them. Thus the total radiation emanating from the surface will be different from that emitted by a black body at the same temperature. That is, the surface will have an "apparent temperature" different from its actual temperature. The concept of the apparent temperature of a surface may be defined in the following manner.

Within an enclosure whose walls are maintained at a fixed temperature T , the black-body radiation is characterized by the quantity $i_o(T, \lambda)$, the power density per unit solid angle of radiation of a specific polarization, with frequencies in the interval Δf . For sufficiently high temperatures and long wavelengths, $i_o(T, \lambda)$ is given by the Rayleigh-Jeans approximation,⁹

$$i_o(T, \lambda) = \frac{kT\Delta f}{\lambda^2} \text{ watts-meter}^{-2}\text{-steradian}^{-1},$$

where k is Boltzmann's constant. For arbitrary temperature and wavelength, the Planck distribution law must be used.

Because of the unique relation between i_o and the temperature T of the enclosure, the radiation itself is also said to have temperature T . Consider now any arbitrary thermal radiation traveling in a specified direction with specified polarization with frequencies in the small interval Δf . Then the apparent temperature of this arbitrary thermal radiation is defined as the temperature of the black-body radiation which has the same power density per unit solid angle. To extend the concept to a surface, it is only necessary to postulate that the apparent temperature of a surface when viewed from some specified direction (with a detector of specified polarization) is equal to the apparent temperature of the radiation emanating from the surface in the prescribed direction with the prescribed polarization.

There are three principal contributions to the apparent temperature of natural terrain surfaces (see Fig. 3).

⁹ J. L. Pawsey and R. N. Bracewell, "Radio Astronomy," Oxford University Press, Oxford Eng., pp. 13-15; 1955.

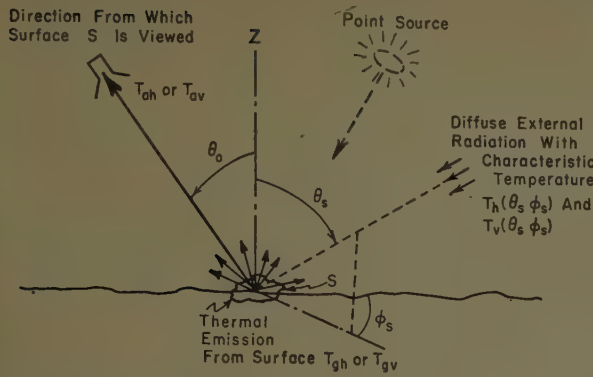


Fig. 3—Contributions to the apparent temperature of a surface.

The first is the thermal radiation emitted by the surface. It is customary to assume that as long as the surface has a well-defined temperature (T_g), its emission coefficient is independent of conditions external to the surface.

The second contribution to the apparent temperature is caused by the reflection of the diffuse thermal radiation (T_h or T_v) impinging on the surface from the upper hemisphere. This diffuse radiation is, in practice, caused primarily by radiation from the earth's atmosphere at microwave wavelengths shorter than 10 cm, and by extraterrestrial (or cosmic) radio noise at wavelengths longer than 100 cm.¹⁰

The third contribution to the apparent surface temperature is caused by radiation from point, or quasi-point sources scattered by the surface. The most prominent of these is the sun.

The power flow from these three principal contributions adds directly to give the total power emanating from the surface. Thus, for conditions under which the Rayleigh-Jeans approximation may be used, the apparent temperature of the surface may also be expressed as a sum of contributing temperatures, *viz*:

$$\begin{aligned} T_{ah} &= T_{gh} + T_{sh} + T_{ph} \\ T_{av} &= T_{gv} + T_{sv} + T_{pv}, \end{aligned} \quad (7)$$

where

$T_{ah}(\theta_o\phi_o)$, $T_{av}(\theta_o\phi_o)$ are the *total* apparent temperatures of a surface for radiation of horizontal or vertical polarization, respectively, when the surface is viewed from the direction $\theta_o\phi_o$ (see Fig. 3).

T_{gh} , T_{gv} are the contributions to the apparent temperature caused solely by the thermal radiation emitted by the surface.

T_{sh} , T_{sv} are the contributions to the apparent temperature caused by the scattering of the diffuse external radiation.

T_{ph} , T_{pv} are the contributions to the apparent temperature caused by the bistatic scattering of radiation from point or quasi-point sources.

The first contribution, that caused by thermal emission from the ground, is easily calculated from the definition (5) of the emission coefficient. Clearly,

$$\begin{aligned} T_{gh}(\theta_o\phi_o) &= e_h(\theta_o\phi_o) T_g \\ T_{gv}(\theta_o\phi_o) &= e_v(\theta_o\phi_o) T_g, \end{aligned} \quad (8)$$

where T_g is the actual temperature of the ground.

The second contribution, caused by the external diffuse radiation may be calculated as follows. Let the apparent temperature of this radiation arriving from the direction $\theta_s\phi_s$ be $T_h(\theta_s\phi_s)$, $T_v(\theta_s\phi_s)$ for the horizontally and vertically polarized components, respectively. Then from the definition of the scattering coefficients, and from the fact that the apparent temperature of radiation is assumed to be proportional to its intensity, it is found that

$$\begin{aligned} T_{sh}(\theta_o\phi_o) &= \frac{1}{4\pi} \int [T_h(\theta_s\phi_s) \gamma_{hh}(s, o) + T_v(\theta_s\phi_s) \gamma_{vh}(s, o)] \\ &\quad \cdot \frac{\cos \theta_s}{\cos \theta_o} d\Omega_s. \end{aligned} \quad (9)$$

From the reciprocity relations, (2), this becomes

$$\begin{aligned} T_{sh}(\theta_o\phi_o) &= \frac{1}{4\pi} \int [T_h(s) \gamma_{hh}(o, s) + T_v(s) \gamma_{hv}(o, s)] d\Omega_s. \end{aligned} \quad (10a)$$

Similarly,

$$\begin{aligned} T_{sv}(\theta_o\phi_o) &= \frac{1}{4\pi} \int [T_v(s) \gamma_{vv}(o, s) + T_h(s) \gamma_{vh}(o, s)] d\Omega_s. \end{aligned} \quad (10b)$$

The third contribution, that caused by point sources, may most easily be calculated by the artifice of replacing the point source by an equivalent source subtending a small solid angle Ω_p , and having an apparent temperature T_p . Then the contribution to the apparent temperature of the surface caused by this quasi-point source is,

$$T_{pi} = (T_p \Omega_p) (4\pi)^{-1} [\gamma_{ii}(o, s) + \gamma_{ij}(o, s)]. \quad (11)$$

For true point sources, only the product $T_p \Omega_p$ is known.⁹

It may be of interest to estimate the contribution of the sun to the apparent surface temperature at microwave frequencies. At K_a band, the apparent temperature of the sun is about 10^4 °K and a typical value of γ is 0.1. Thus T_{sh} caused by the sun is about 0.01°K. It seems likely that, unless great accuracy is required, the effects of the sun on apparent temperature may be ignored for diffuse surfaces.

¹⁰ P. D. Strum, "Considerations in high sensitivity microwave radiometry," *PROC. IRE*, vol. 46, pp. 43-53; January, 1958.

Two qualifications to the above must be mentioned explicitly. First, all the matter in and below the surface must be at a uniform temperature. Secondly, the apparent temperatures calculated above apply only near enough to the surface so that there is no appreciable attenuation between surface and observer at the frequency of interest. If this condition is not fulfilled, the attenuation and thermal radiation caused by the intervening atmosphere must also be taken into account.

APPLICATION TO THE APPARENT TEMPERATURES OF SOME NATURAL SURFACES

In the previous sections, general formulas for apparent surface temperature were derived. They will now be applied to two surfaces for which some measurements are available,¹ and which represent fairly well the two extremes of surface condition met in practice, *viz.*, the practically smooth surface, and the very rough surface.

The first surface of interest is an asphalt roadway. Theoretical considerations^{3,11} confirmed by scattering measurements^{4,12} show that the scattering from such a slightly rough surface is predominantly specular, with a reflection coefficient not much different from that for a flat surface having the same dielectric constant. The diffuse part of the scattering pattern, though of vital importance for predicting radar return, contains only a few per cent as much power as does the specular part. Thus for an approximate calculation of apparent temperature, the diffuse scattering may be neglected and the surface treated as perfectly flat. The emission and reflection coefficients of such a surface have long been known.⁶ However, from the more general point of view taken here, one has merely to insert the δ -function representation of the specular scattering into (7) to obtain the apparent temperatures,

$$\begin{aligned} T_{ah}(\theta_o\phi_o) &= [1 - |R_h(\theta_o)|^2]T_a \\ &\quad + |R_h(\theta_o)|^2T_h(\theta_o, \phi_o + \pi) \\ T_{av}(\theta_o\phi_o) &= [1 - |R_v(\theta_o)|^2]T_a \\ &\quad + |R_v(\theta_o)|^2T_v(\theta_o, \phi_o + \pi). \end{aligned} \quad (12)$$

Here R_h and R_v are the Fresnel reflection coefficients.⁵ It may be mentioned that for the general surface, (12) also gives the correct form for the contribution to apparent temperature caused by the specular part of the scattering if the $|R_i|^2$ are replaced by the appropriate specular reflection coefficients.

In order to use (12), it is necessary to estimate the apparent "sky" temperatures $T_h(\theta)$ and $T_v(\theta)$. At 4.3 mm wavelength, these are caused primarily by radiation from the earth's atmosphere. A convenient formula, based on a simple model which represents the atmos-

phere as a uniform horizontal layer is

$$T_h(\theta_o) = T_v(\theta_o) = T_a(1 - r^{\sec\theta_o}), \quad (13)$$

where r is the fraction of power of the given wavelength transmitted through the atmosphere at normal incidence and T_a is the temperature of the model atmosphere. Taking $r=0.63$, a value appropriate to the dry air of Texas, the predictions of (13) are in fair agreement with the sky temperatures measured by Straiton. With this expression for sky temperature, and assuming that the surface has a dielectric constant of 4.0, the predictions of (12) are shown in Fig. 4. The measurements of Straiton, Britt and Tolbert¹ on an asphalt surface are included for comparison. It will be seen that, apart from a systematic error of about 10°K, both the polarization dependence and the angular dependence of the measured temperatures are fairly well represented by the model.

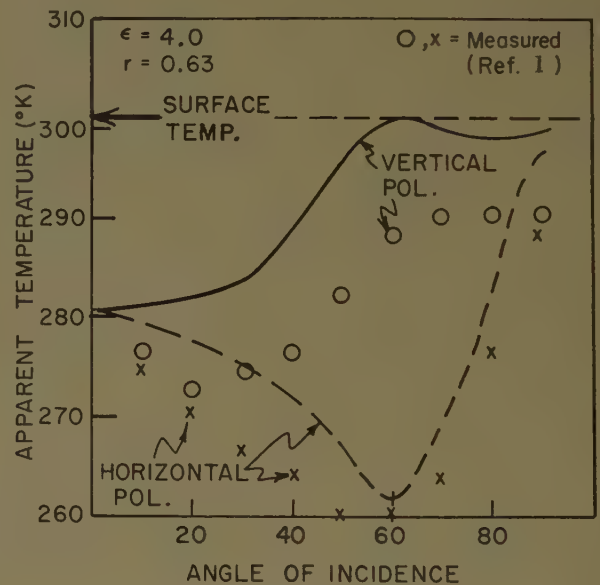


Fig. 4—Measured and calculated apparent temperature of an asphalt surface.

A second surface measured by Straiton *et al.*, for which the differential scattering coefficients may be estimated, was one composed of grass and weeds. For such a surface (at 4.3 mm wavelength) the scattering is predominantly diffuse with no specular reflection except at grazing incidence. Thus, for an approximate calculation of the apparent temperature of such a surface, one may consider only the diffuse scattering. Radar return measurements by Taylor⁴ show that the back-scattering $\gamma(o, o)$ tends to be independent of incidence angle, and measurements reported by Campbell¹³ indicate that the cross-polarized return is somewhat smaller than, but similar in character to, the direct re-

¹¹ W. H. Peake, "Theory of radar return from terrain," 1959 IRE NATIONAL CONVENTION RECORD, pt. 1, pp. 27-41.

¹² M. P. Bachynski, "Microwave propagation over rough surfaces," *RCA Rev.*, vol. 20, pp. 308-335; June, 1959.

¹³ J. P. Campbell, "Back scattering characteristics of land and sea at X-band," *Proc. Natl. Conf. on Airborne Electronics*, Dayton, Ohio, p. 107; May, 1958.

turn. A simple heuristic choice for $\gamma(o, s)$ consistent with both these measurements and the reciprocity principle is

$$\gamma_{vv}(o, s) + \gamma_{vh}(o, s) = \gamma_o(\cos \theta_o + \cos \theta_s)/(2 \cos \theta_o), \quad (14)$$

where γ_o is a constant. A similar expression should hold for horizontal polarization. This special choice for $\gamma(o, s)$, together with the sky temperature given by (13) may be substituted into (7) to obtain the apparent temperature. The results are shown in Fig. 5, in which ΔT , the difference between the actual and apparent temperature of the surface, is given as a function of incidence angle, together with the measured values.¹ The lower curve refers to dry grass and the upper one to damp grass; Taylor's measurements of back-scattering from dry and damp grass (at 8.6 mm) indicate that appropriate values of the constant γ_o are 0.1 and 0.2 respectively. Again, there is reasonably good agreement between measured and calculated values except near grazing incidence.

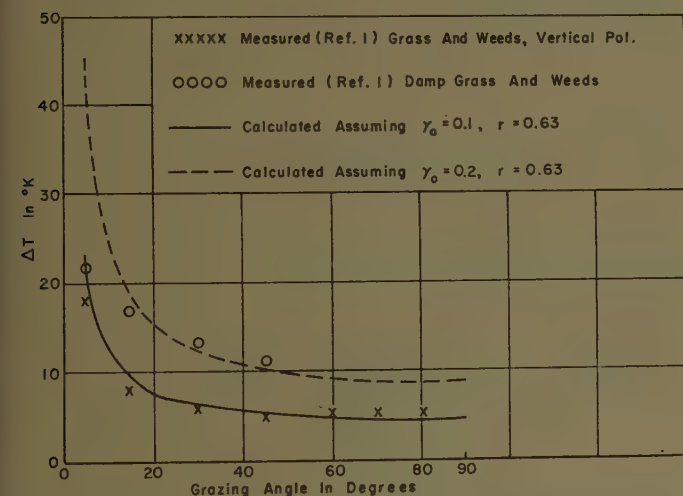


Fig. 5—Measured and calculated values of the difference between actual surface temperature and apparent temperature of a surface of grass and weeds.

Both of the above examples suffer from the fact that the complete scattering pattern of the surfaces (including both the specular and the diffuse part) were not known. Nevertheless they may be useful in illustrating the effects of different kinds of terrain on antenna temperature. However, an accurate calculation of antenna temperature must depend on much more detailed knowledge of surface scattering properties than is now available. This is particularly important in the case of tropospheric scatter antennas, where, because the main beam lies along the horizon, the details of the transition to specular reflection at grazing angles may have a significant effect on antenna temperature. A more complete discussion of the scattering properties of terrain is given in another report.¹⁴

CONCLUSION

It has been shown that the "bistatic" or "differential" scattering coefficients γ_{hh} , etc., may be used to determine a number of properties of the interaction of electromagnetic radiation with natural surfaces, in particular the albedo and the absorption coefficient. By making use of the reciprocity relations satisfied by the scattering coefficients, it is then possible to derive Kirchhoff's radiation law in a form which accounts for the angular dependence and polarization dependence of the emission and absorption coefficients and predicts them in terms of the γ 's. Thus it has been shown that once the γ 's are known, both the scattering and the radiometric properties of the surface may be found in terms of them. The results have been used to predict the apparent temperatures of an asphalt- and a grass-covered surface, and are in fair agreement with the measured apparent temperatures.

¹⁴ W. H. Peake, "Interaction of Electromagnetic Waves with Some Natural Surfaces," Ph.D. dissertation, The Ohio State University, Columbus; 1959.

Electromagnetic Scattering by High-Density Meteor Trails*

H. BRYSK†

Summary—The discussion first concerns itself with establishing limits on the validity of the low-density approximation. These depend not only on the electron line density (as has sometimes been loosely stated) but also on the wavelength of observation and on the altitude of the trail. Next, a model is developed for scattering by a supercritical density distribution of electrons, based on the idea that the process can still be viewed as a superposition of individual Compton effects, but with the wave incident on the electron attenuated because of refraction (in analogy to the skin effect). Results of this model are compared with those obtained by the usual approach of replacing the electron distribution by a metallic scatterer whose surface is the critical density contour. Some calculations with non-Gaussian electron distributions help to clarify the physical interpretation.

I. PHASE ERROR IN INDIVIDUAL SCATTERER MODEL

SCATTERING by a meteor trail of low electron density is usually treated by considering each electron to be subject to the incident field and to scatter independently (according to the Thomson cross section), the contributions from the electrons being added up coherently.¹

For a moderately higher density (still below the critical density everywhere), with an electron collision frequency well below the frequency of the radiation, the model is disturbed by the deviation from unity of the index of refraction of the inner region (which alters the phase relations between electrons). As an indication of when this effect can become significant, the phase error made in neglecting the variation of the index of refraction is computed for a ray coming in from infinity to the axis of a cylindrical Gaussian distribution. When this phase error is small, the individual scatterer model is good; as it becomes large, the model breaks down.

The index of refraction n is given by

$$n^2 = 1 - (4\pi N r_0 / k^2), \quad (1)$$

where

N = electron density,

$r_0 = e^2 / mc^2$ = classical electron radius,

$k = 2\pi / \text{wavelength}$.

The electron distribution in a meteor trail is approximated locally by a cylindrical Gaussian;

$$N(r) = (\alpha / \pi a^2) \exp(-r^2 / a^2). \quad (2)$$

The phase error ϕ is then

$$\phi = k \csc \theta \int_0^\infty (1 - n) dr, \quad (3)$$

where θ is the angle between the trail axis and the ray (under the stipulated conditions, the bending of the ray can be neglected). From (1) and (2)

$$n^2(r) = 1 - (4\alpha r_0 / k^2 a^2) \exp(-r^2 / a^2) \quad (4)$$

and the minimum value of n is that at $r=0$;

$$n^2(0) = 1 - (4\alpha r_0 / k^2 a^2). \quad (5)$$

The expression in (3) can be reasonably well bounded by noting that

$$1 - n = (1 - n^2) / (1 + n). \quad (6)$$

Since

$$1 \geq n \geq n(0), \quad (7)$$

$$2 \geq 1 + n \geq 1 + n(0); \quad (8)$$

and

$$\begin{aligned} [1 + n(0)]^{-1} k \csc \theta \int_0^\infty (1 - n^2) dr \\ \geq \phi \geq 2^{-1} k \csc \theta \int_0^\infty (1 - n^2) dr. \end{aligned} \quad (9)$$

The integral is simple,

$$\begin{aligned} \int_0^\infty (1 - n^2) dr \\ = (4\alpha r_0 / k^2 a^2) \int_0^\infty dr \exp(-r^2 / a^2) = 2\pi^{1/2} \alpha r_0 / k^2 a, \end{aligned} \quad (10)$$

so that

$$\begin{aligned} 2\pi^{1/2} \alpha r_0 \csc \theta / ka \{1 + [1 - (4\alpha r_0 / k^2 a^2)]^{1/2}\} \\ \geq \phi \geq \pi^{1/2} \alpha r_0 \csc \theta / ka. \end{aligned} \quad (11)$$

It is convenient to refer to the characteristic dimensionless constants

$$B = 2(\alpha r_0)^{1/2} \quad (12)$$

$$K = ka. \quad (13)$$

In terms of these, the last equation reads

$$\begin{aligned} 2^{-1} \pi^{1/2} (B^2 / K) \csc \theta \{1 + [1 - (B/K)^2]^{1/2}\}^{-1} \\ \geq \phi \geq 4^{-1} \pi^{1/2} (B^2 / K) \csc \theta. \end{aligned} \quad (14)$$

For $B \ll K$, the phase error ϕ is exactly determined. At worst, for $B = K$ (critical density reached on axis), there is a factor of 2 uncertainty.

* The research reported in this paper has been sponsored by the Electronics Research Directorate, AF Cambridge Research Center, Air Res. and Dev. Command, USAF, under Contract AF 19(604)-4993.

† Physics Dept., Adelphi College, Garden City, N. Y. Formerly at Radiation Laboratory, University of Michigan, Ann Arbor, Mich.

¹ H. Brysk, "Electromagnetic scattering by low density meteor trails," *J. Geophys. Res.*, vol. 63, pp. 693-716; December, 1958.

The definition of what constitutes a small phase error is somewhat arbitrary. The simplest reasonable choice from (14) is

$$(B^2/K) \leq 1; \quad (15)$$

(for $\theta=0$, this leads to $\phi \leq 25^\circ$ from the right-hand estimate). At the same time, if the trail is to be of subcritical density everywhere

$$B \leq K. \quad (16)$$

Hence, the rough criterion of applicability of the underdense model is that

$$B \leq K, \quad B < 1 \quad (17)$$

$$B^2 \leq K, \quad B > 1. \quad (18)$$

Fig. 1 exhibits the limiting value of the electron line density as a function of the wavelength, using for a the initial width computed by Öpik² with densities at four altitudes obtained from Watanabe,³ for a meteor velocity of 40 km/second. This set of curves is to be contrasted with the frequently made unqualified statement⁴ that the transition from the underdense to the overdense case occurs at $\alpha = 10^{12}$ electrons/cm.

II. ATTENUATION FACTOR DUE TO PHASE VARIATION ON CRITICAL DENSITY CONTOUR

Consider a plane perpendicular to the trail axis. If there is a critical density region, the intersection of this plane with the critical density contour will be a circle. The phase variation of rays reaching this circle will next be investigated. For simplicity, the rays will be assumed parallel. (This is fully justified, comparing the radius of the critical density region \bar{R} with the source size and range, although it would be incorrect over the length of the trail.)

The phase of a ray coming in from infinity and reaching the critical density contour is, in general,

$$\phi = k \csc \theta \int_{\bar{R} \cos \beta}^{\infty} n(\beta) dz, \quad (19)$$

where β is the azimuthal angle.

For a ray whose extension would hit the axis of the trail,

$$\phi_0 = k \csc \theta \int_{\bar{R}}^{\infty} n_0 dz. \quad (20)$$

The phase difference between the rays is

$$\phi - \phi_0 = k \csc \theta \left\{ \int_{\bar{R} \cos \beta}^{\bar{R}} n dz - \int_{\bar{R}}^{\infty} [n_0 - n] dz \right\}. \quad (21)$$

² E. J. Öpik, "Physics of Meteor Flight in the Atmosphere," Interscience Publishers, Inc., New York, N. Y.; 1958.

³ K. Watanabe, "Ultraviolet absorption processes in the upper atmosphere," *Advances in Geophys.*, vol. 5, pp. 153-221; 1958.

⁴ L. A. Manning and V. R. Eshleman, "Meteors in the Ionosphere," *Proc. IRE*, vol. 47, pp. 186-199; February, 1959.

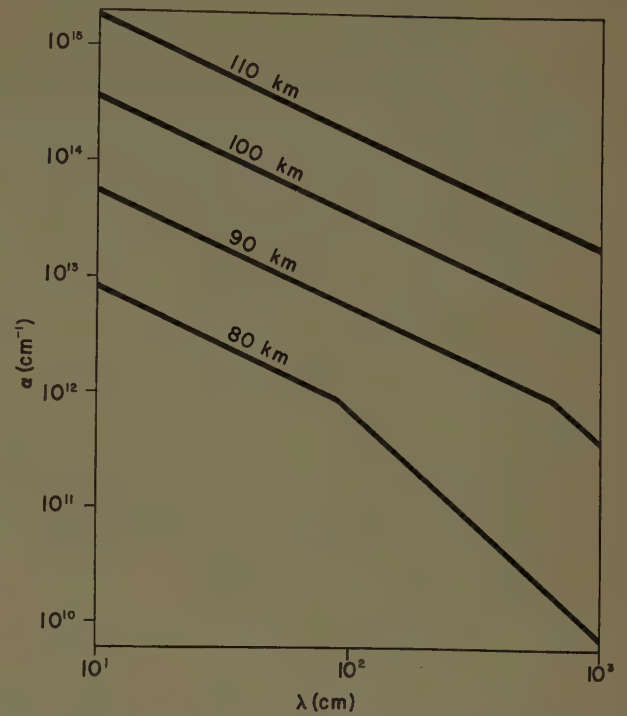


Fig. 1—Maximum electron line density for which independent scatterer model assumptions are valid.

The index of refraction is given by

$$n^2 = 1 - (B/K)^2 \exp[-r^2/a^2] \\ = 1 - (B/K)^2 \exp[-(z^2 + \bar{R}^2 \sin^2 \beta)/a^2]. \quad (22)$$

Since the index of refraction is zero at the critical density contour, $(B/K)^2$ and R are related, so that (22) can be rewritten more concisely as

$$n^2 = 1 - \exp[(\bar{R}^2 \cos^2 \beta - z^2)/a^2]. \quad (23)$$

To evaluate the integrals approximately, an inequality like (7) is again resorted to, after expressing

$$n = 1 - (1 - n^2)/(1 + n) \quad (24)$$

$$n_0 - n = (n_0^2 - n^2)/(n_0 + n), \quad (25)$$

by replacing the denominators by estimated mean values. Thus,

$$\phi - \phi_0 = k \csc \theta \left\{ \bar{R}(1 - \cos \beta) - \overline{(1 + n)}^{-1} \right. \\ \cdot \exp(\bar{R}^2 \cos^2 \beta/a^2) 2^{-1} \pi^{1/2} a [\Phi(\bar{R}/a) - \Phi(\bar{R} \cos \beta/a)] \\ \left. + \overline{(n_0 + n)}^{-1} [\exp(\bar{R}^2/a^2) - \exp(\bar{R}^2 \cos^2 \beta/a^2)] 2^{-1} \pi^{1/2} a \right. \\ \left. \cdot [1 - \Phi(\bar{R}/a)] \right\}. \quad (26)$$

In the asymptotic limit for $\bar{R} \cos \beta \gg a$, the last two terms in (26) are small to order $(a/\bar{R})^2$ compared with the first, so that (26) can be reduced approximately to

$$\phi - \phi_0 = k \bar{R} \csc \theta (1 - \cos \beta). \quad (27)$$

As (a/\bar{R}) increases, a less tidy situation ensues. For a reasonable indication of the effect of phase variation (accepting a possible error in phase of as much as a factor of 2), (27) will be used.

Within the critical density contour the wave suffers no further phase change. Hence, if the electron distribution has cylindrical symmetry, contributions for all values of β are equiprobable. Due to these phase differences, there is an attenuation factor f_p at a given point along the axis, given for a two-way traversal by

$$f_p = (2\pi)^{-1} \int_{-\pi}^{\pi} d\beta \exp [-2ik\bar{R} \csc \theta \cos \beta] \\ = J_0(2k\bar{R} \csc \theta). \quad (28)$$

For large argument, the asymptotic expression yields

$$f_p = (\pi k \bar{R} \csc \theta)^{-1/2} \cos [2k\bar{R} \csc \theta - (\pi/4)]. \quad (29)$$

Averaging out the oscillatory factor (*i.e.*, averaging over time during the expansion of the trail, or simply recognizing that the critical density contour is less than perfectly sharp), (29) reduces to

$$f_p = (2\pi k \bar{R} \csc \theta)^{-1/2}. \quad (30)$$

III. ATTENUATION IN SUPERCRITICAL REGION

Scattering by an overdense region is usually handled by replacing the region by a perfect conductor whose surface is defined by the critical density contour. Aside from problems of computation of the scattering by the conductor, there are intrinsic objections to this model. An ionized region can maintain a transverse current, while a conductor cannot. Within a metal, the real and imaginary parts of the index of refraction are equal in magnitude, whereas within the overdense ionized region the real part is zero (so that phase properties are entirely different). Furthermore, the critical density contour is not a perfect reflector. There is penetration into the overdense region with an attenuation factor, analogous to the skin effect for a metal. The electron density usually increases toward the center of the overdense region, so that there are the counteracting trends of the skin effect tending to reduce the scattering from deep inside per electron while the density distribution indicates an increase in the number of scattering centers.

The approach attempted here is a mongrel model. The scattering is considered to consist of three distinct regimes: the incoming wave, the Compton process itself, and the outgoing wave. The Compton process is treated as an individual particle effect, just as for the underdense case. The waves, on the other hand, are handled from a ray-tracing viewpoint. The amplitude of the wave being scattered by the electron is taken to be the incident amplitude reduced by the skin-effect attenuation, the latter being computed along the shortest optical path from the critical density contour to the electron. The amplitude of the scattered wave is reduced by the same factor in coming out of the overdense region. It should be noted that the skin effect represents a reduction of amplitude due to refraction away from a region rather than absorption in the region. The discussion assumes that the frequency of the electro-

magnetic radiation is sufficiently higher than the electron collision frequency that absorption can be neglected.

The results will be expressed as the ratio of the scattering amplitude obtained from the overdense region to that which would be obtained from the same number of electrons treated as individual scatterers without phase differences; *i.e.*, in effect, the reduction in effectiveness of scattering due to the denseness.

In particular, for a cylindrically symmetrical distribution of electrons the attenuation factor f' will be given by

$$f' = \left\{ \int_0^{\bar{R}} r dr N(r) \cdot \exp \left\{ -2k \int_0^{\bar{R}} \bar{n}(r') dr' \right\} \right\} / \int_0^{\bar{R}} r dr N(r) \quad (31)$$

where \bar{n} denotes the absolute value of n .

For the Gaussian distribution, with the further notation

$$x = r/a, \quad X = \bar{R}/a, \quad (32)$$

the exponent in (31) becomes

$$2k \int_r^{\bar{R}} \bar{n}(r') dr' = 2K \int_x^X [(B/K)^2 \exp(-t^2) - 1]^{1/2} dt \\ = 2B \int_x^X [\exp(-t^2) - \exp(-X^2)]^{1/2} dt. \quad (33)$$

The denominator is simply

$$\int_0^{\bar{R}} r dr N \\ = (\alpha/\pi) \int_x^X x dx \exp(-x^2) = (\alpha/2\pi) [1 - \exp(-X^2)]. \quad (34)$$

Thus, the attenuation factor is

$$f' = 2[1 - \exp(-X^2)]^{-1} \int_0^X x dx \exp(-x^2) \\ \cdot \exp \left\{ -2B \int_x^X [\exp(-t^2) - \exp(-X^2)]^{1/2} dt \right\}. \quad (35)$$

The indicated integrations cannot be done analytically. Two special cases will be studied below.

A. Very High Density

If the electron line density parameter B is much larger than the critical-density-region size parameter K , (35) simplifies a bit on going to the limit $X \rightarrow \infty$. (This limit can be used provided X is greater than about 3.) The integral in the exponent becomes

$$\int_x^X [\exp(-t^2) - \exp(-X^2)]^{1/2} dt \\ \rightarrow \int_x^\infty \exp(-t^2/2) dt = (\pi/2)^{1/2} \operatorname{erfc}(x/2^{1/2}) \quad (36)$$

so that (35) reduces to

$$f' = 2 \int_0^\infty x dx \exp(-x^2) \exp[-(2\pi)^{1/2} B \operatorname{erfc}(x/2^{1/2})]. \quad (37)$$

In this last form, f' depends only on B ; i.e., only on α . Thus for a sufficiently dense region, the attenuation depends only on the line density of electrons, and not on the wavelength of the radiation or the width of the distribution. These quantities do enter (in the combination ka) in the determination of what constitutes a "sufficiently dense" region. Fig. 2 exhibits f' , from (37), as a function of α ; on a log-log plot, the curve is very nearly a straight line.

The scattering amplitude of the meteor trail is proportional to $\alpha f'$ and hence to $B^2 f'$. In Fig. 3, $B^2 f'$ has been plotted against B . The curve is fitted by

$$B^2 f' = 0.91 \ln B \quad (38)$$

apart from the small- B end [where (37) is not a good approximation anyhow].

For an intuitive grasp of the characteristics of the scattering, it is instructive to compare (38) with the scattering amplitude obtained for two simpler electron distributions—a uniform cylinder and two coaxial uniform cylinders—with the same line density and in the high density limit.

For the uniform cylinder, N is a constant, hence \bar{n} is also a constant, and (31) reduces to

$$\begin{aligned} f' &= \left\{ N \int_0^{\bar{R}} r dr \exp \left[-2k\bar{n} \int_r^{\bar{R}} dr' \right] \right\} / N \int_0^{\bar{R}} r dr \\ &= (2/\bar{R}^2) \{ \exp[-2k\bar{n}\bar{R}] \} \int_0^{\bar{R}} r dr \exp[2k\bar{n}r] \\ &= (k\bar{n}\bar{R})^{-1} \{ 1 - (2k\bar{n}\bar{R})^{-1} [1 - \exp(-2k\bar{n}\bar{R})] \}. \end{aligned} \quad (39)$$

For a large electron density, the second term in (1) predominates, so that

$$k\bar{n} = (4\pi N r_0)^{1/2}. \quad (40)$$

The electron line density α is

$$\alpha = \int_0^{2\pi} d\phi \int_0^{\bar{R}} r dr N = \pi \bar{R}^2 N \quad (41)$$

so that $k\bar{n}\bar{R}$ reduces to

$$k\bar{n}\bar{R} = 2(\alpha r_0)^{1/2} = B. \quad (42)$$

For large B , the attenuation factor is then simply

$$f' = B^{-1} \quad (43)$$

and the scattering amplitude is

$$B^2 f' = B. \quad (44)$$

For two coaxial uniform cylinders, the procedure is analogous except that the r -integral is split into two portions (0 to \bar{R}' and \bar{R}' to \bar{R}) for each of which N and \bar{n} are constants (denoted by subscripts 1 and 2, respectively).

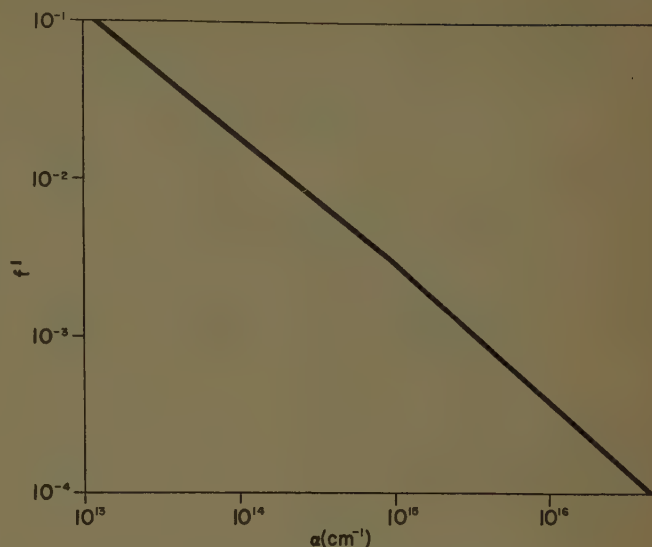


Fig. 2—Attenuation factor for very high density.

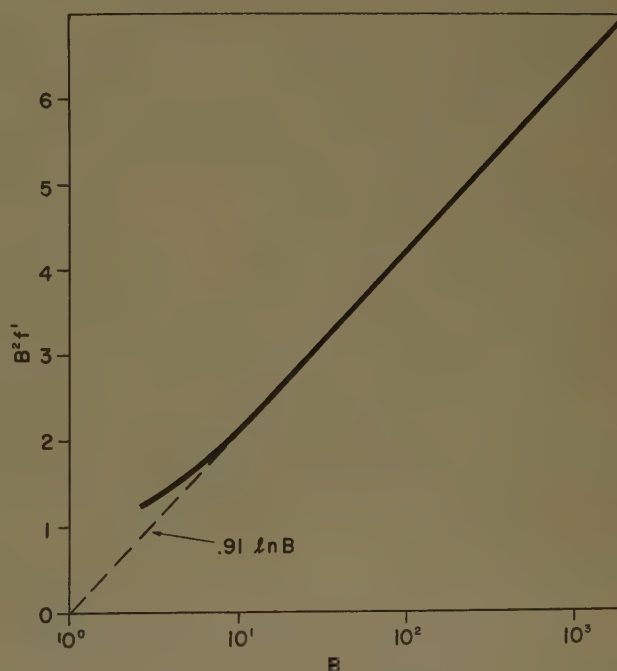


Fig. 3—Scattering amplitude for very high density.

The denominator of (31) is always $(\alpha/2\pi)$, so

$$\begin{aligned} \alpha f' / 2\pi &= N_1 \int_0^{\bar{R}'} r dr \exp \left\{ -2k \left[\bar{n}_1 \int_r^{\bar{R}'} dr' + \bar{n}_2 \int_{\bar{R}'}^{\bar{R}} dr' \right] \right\} \\ &\quad + N_2 \int_{\bar{R}'}^{\bar{R}} r dr \exp \left[-2k\bar{n}_2 \int_r^{\bar{R}} dr' \right] \\ &= (N_1 \bar{R}' / 2k\bar{n}_1) \{ 1 - (2k\bar{n}_1 \bar{R}')^{-1} [1 - \exp(-2k\bar{n}_1 \bar{R}')] \} \\ &\quad \cdot \exp[-2k\bar{n}_2(\bar{R} - \bar{R}')] \\ &\quad + (N_2 \bar{R} / 2k\bar{n}_2) [1 - (2k\bar{n}_2 \bar{R})^{-1}] - (N_2 \bar{R}' / 2k\bar{n}_2) \\ &\quad \cdot [1 - (2k\bar{n}_2 \bar{R}')^{-1}] \exp[-2k\bar{n}_2(\bar{R} - \bar{R}')]. \end{aligned} \quad (45)$$

In the limit of large density and extent for both regions [*i.e.*, for (40) valid and the negative exponentials on the right-hand side of (45) all small], there results

$$B^2 f' = (4\pi r_0 N_2)^{1/2} \bar{R}. \quad (46)$$

Note that

$$\begin{aligned} \alpha &= 2\pi \left[N_1 \int_0^{\bar{R}'} r dr + N_2 \int_{\bar{R}'}^{\bar{R}} r dr \right] \\ &= \pi [N_1 \bar{R}'^2 + N_2 (\bar{R}^2 - \bar{R}'^2)] \\ &= \pi N_2 \bar{R}^2 + \pi (N_1 - N_2) \bar{R}'^2. \end{aligned} \quad (47)$$

If the inner region has the higher electron density, there is thus a lower f' (for a given α) than in the uniform cylinder case.

For definiteness, the coaxial cylinder case will be specialized by making the proviso that each of the two regions contain the number of electrons that would be present in the corresponding part of the Gaussian. Then

$$\begin{aligned} \int_{\bar{R}'}^{\bar{R}} r dr N &= N_2 (\bar{R}^2 - \bar{R}'^2)/2 \\ &= (\alpha/2\pi) [\exp(-\bar{R}'^2/a^2) - \exp(-\bar{R}^2/a^2)]. \end{aligned} \quad (48)$$

It was shown above that

$$\exp(-\bar{R}^2/a^2) = (K/B)^2. \quad (49)$$

For convenience, write

$$\bar{R}'^2 = p \bar{R}^2 \quad (50a)$$

where

$$0 < p < 1, \quad (50b)$$

so that

$$\exp(-\bar{R}'^2/a^2) = (K/B)^{2p}. \quad (51)$$

The value of N_2 can then be expressed simply by

$$(\pi/\alpha) N_2 (1-p) \bar{R}^2 = (K/B)^{2p} - (K/B)^2 \quad (52)$$

so that (46) becomes

$$B^2 f' = B(1-p)^{-1/2} [(K/B)^{2p} - (K/B)^2]^{1/2}. \quad (53)$$

As the discussion applies to high densities ($B \gg K$), only the first term in the bracket need be retained since $p < 1$ (with some care that p not be too close to 1). Thus

$$B^2 f' = (1-p)^{-1/2} K^p B^{1-p}. \quad (54)$$

Comparing now the high-density limits of the three cases, we find that:

1) for a given line density, the scattering amplitude is less for the two-region case (with inner region more dense) than for the uniform cylinder case, and still less for the Gaussian;

2) the scattering amplitude for the two-region case is in fact, according to (46), just that which would occur if both regions had the density of the outer one;

3) the scattering amplitude varies as the square root of the line density in the uniform region case, as a smaller positive power of the line density in the two-

region case, and as its logarithm (still slower) in the Gaussian case.

The implication of these observations is that, for a region of radially decreasing high electron density, the scattering characteristics are predominantly determined by the outer portions of the supercritical density region (the core not being sufficiently penetrated by the radiation). The result from increasing the line density of electrons is primarily to push the effective scattering region outward, rather than to increase its density. This does result in an increase in the scattering because the surface area is increased (hence more electrons are accessible).

B. Critical Density Contour of Maximum Width

The electron distribution spreads out in time due to diffusion, and a increases. From (4) it can be deduced that the radius of the critical density region is given by

$$\bar{R}^2 = a^2 \ln(4\alpha r_0/k^2 a^2). \quad (55)$$

The maximum value that it can attain (as a function of a —*i.e.*, of time) is given by

$$\partial \bar{R}^2 / \partial a^2 = \ln(4\alpha r_0/k^2 a^2) - 1 = 0 \quad (56)$$

which means that

$$X = 1. \quad (57)$$

Heuristically, the widest critical density contour might be expected to yield the largest scattering return from the supercritical region, because it corresponds to the greatest number of electrons being exposed to an unattenuated incident field. This argument would be much weakened if it should turn out that f' (viewed as a function of X) has resonance-type oscillations.

Ideally, the maximum value of f' should be obtained by setting $\partial f' / \partial X = 0$, where f' is given by (35). Unfortunately, this is impractical because $\partial f' / \partial X$ includes a term in f' and also a term involving an integral like f' with an additional integral as a factor in the integrand, as well as a term independent of the f' -integral; the equation $\partial f' / \partial X = 0$ cannot be solved unless both f' and the somewhat more complicated companion integral are known as a function of X . Hence, there is no direct way of determining the maximum value of f' short of actually computing f' as a function of X . Since this must then be repeated for each B of interest, the computational effort required is quite large.

In what follows, it will be assumed that the maximum value of f' is indeed attained for $X = 1$. This is also the assumption in the metallic scatterer approximation, so there will be a direct comparison of results for the same configuration. The resultant special case of (35) is

$$\begin{aligned} f' &= 2[1 - e^{-1}]^{-1} \int_0^1 x dx \exp(-x^2) \\ &\cdot \exp \left\{ -2B \int_0^1 [\exp(-t^2) - e^{-1}]^{1/2} dt \right\}. \end{aligned} \quad (58)$$

This is plotted in Fig. 4. The curve is well fitted by

$$f' = 1.05B^{-.69}. \quad (59)$$

So far, the "skin effect" attenuation and the phase relations around the critical density contour have been considered. There remains to examine the phase change along the trail axis due to the deviation of the index of refraction from unity. This consists of two contributions: the phase error in reaching the critical density contour (evaluated in Section I), plus the error incurred in including the region inside the critical density region in the optical path

$$\phi' = k\bar{R} \csc \theta. \quad (60)$$

During the growing phase (case A), the contribution of (14) dominates over that of (60) and leads to oscillations superposed on the scattering integrand. Near the maximum expansion of the critical density contour (case B), on the other hand, the two contributions are comparable and vary in opposite directions, so that the longitudinal phase variation is very slow.

Accordingly, for the maximum critical density contour the attenuation factor, instead of the exponential of Brysk¹ is approximately the product of the phase reduction factor of (30) and the "skin effect" attenuation factor of (59). In (30), the relations for this maximum lead to

$$k\bar{R} = ka = K = e^{-1/2}B \quad (61)$$

so that

$$f = (2\pi e^{-1/2}B \csc \theta)^{-1/2} 1.05B^{-.69} = .52 \sin^{1/2} \theta B^{-1.19}. \quad (62)$$

The maximum return is now given by

$$S = .076 \csc \theta PGG'\lambda^3(\alpha r_0)^{.81}(\hat{\theta} \cdot \hat{\theta}')^2/16\pi^2 RR'(R + R') \quad (63)$$

or monostatically

$$S = .038 PGG'\lambda^3(\alpha r_0)^{.81}(\hat{\theta} \cdot \hat{\theta}')^2/16\pi^2 R^3. \quad (64)$$

This result is to be compared with that obtained by assuming the critical density region to scatter the electromagnetic radiation like a metallic cylinder, the return from the latter being computed by geometrical optics, *i.e.*, in the limit $\bar{R} \gg \lambda$.⁵ The correct geometrical "cross section" for broadside backscattering from a metallic cylinder is

$$\sigma = \pi \bar{R} R. \quad (65)$$

This is a factor of two less than that quoted by Greenhow,⁵ and it leads to

$$S = .048 PGG'\lambda^3(\alpha r_0)^{1/2}(\hat{\theta} \cdot \hat{\theta}')^2/16\pi^2 \cdot R^3. \quad (66)$$

Fig. 5 compares the maximum return from an overdense trail as obtained by the present "skin effect" model with that obtained by the "metallic cylinder" model.

⁵ J. S. Greenhow, "Characteristics of radio echoes from meteor trails: III, the behaviour of the electron trails after formation," *Proc. Phys. Soc., B*, vol. 65, pp. 169-181; March, 1952.

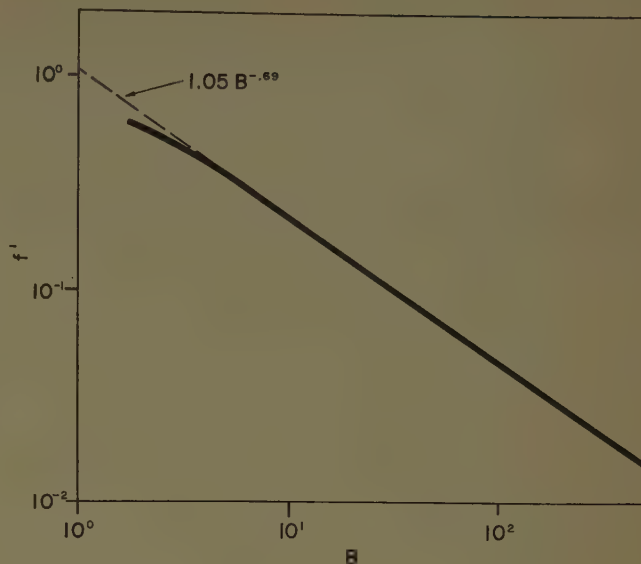


Fig. 4—Attenuation factor for critical density contour of maximum extent.

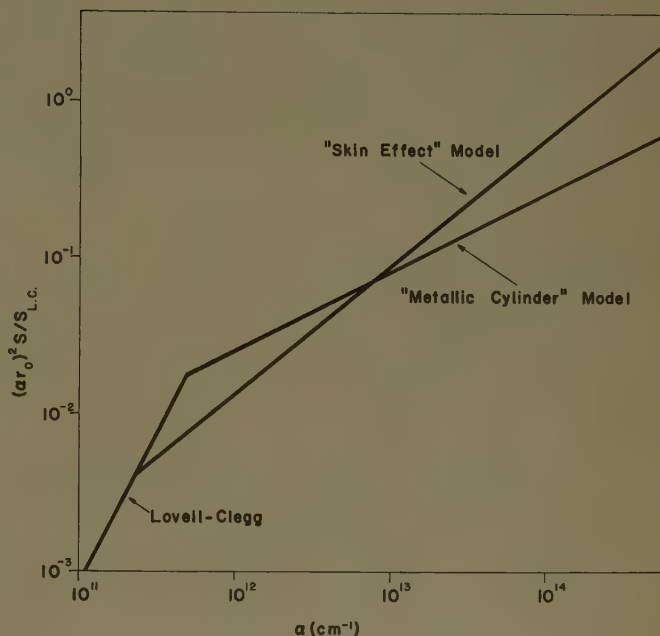


Fig. 5—Comparison of "skin effect" model with "metallic cylinder" model.

model. The Lovell-Clegg low-density result is also shown for orientation. For clarity of display exclusively, the overdense and underdense trail curves have been extended to meet. The temptation to bridge the transition by fairing in from one curve to the other should be resisted; the intermediate region very probably does not behave that simply. The two overdense trail curves cross for relatively low α ($\sim 7 \times 10^{13} \text{ cm}^{-1}$). On the low- α side, they yield essentially undistinguishable predictions (especially taking into account the pile-up of theoretical errors in the transition region). On the high- α side, the "skin effect" model yields an increasingly larger result, as expected intuitively.

A feature deserving of comment is the specular nature of the return. In this respect, the present model leads to the same variation as the underdense trail model. The metallic cylinder model, of course, shows a strong aspect dependence. On the other hand, a loss of specularity as the electron line density increases has been reported experimentally. In so far as there is such a loss of specularity in the relatively early history of the trail

(*i.e.*, before enough time has elapsed for some form of turbulence to be invoked), it would appear that the basic assumption of a uniform line density of ionization underlying both models must be abandoned; and that, although a uniform line density can be satisfactorily assumed for underdense trails, it is an essential feature of scattering by overdense trails that the line density is markedly nonuniform.

Electromagnetic Properties of High-Temperature Air

M. P. BACHYNSKI†, T. W. JOHNSTON†, AND I. P. SHKAROFSKY†

Summary—This paper concerns the attenuation and phase characteristics of plasmas and, in particular, the electromagnetic properties of high-temperature air. It is shown that by a suitable normalization of the parameters the e - m properties of plasmas may be universally represented in convenient form in either the complex dielectric coefficient plane or the complex propagation constant plane.

INTRODUCTION

BECAUSE of ionization, air at high temperatures contains an appreciable number of free electrons and ions. Under these conditions, the medium may be described as a plasma, *i.e.*, a gas containing charged particles in a sufficient quantity to seriously alter the physical properties of the gas. One of the properties of air markedly affected by the presence of the electrons and ions is the propagation of electromagnetic waves in such a medium. This interaction of electromagnetic waves with plasmas is of current interest in connection with diagnostic techniques, space communications, and re-entry problems.

This paper is concerned with the electromagnetic characteristics of plasmas and, in particular, those of high-temperature air. It is shown that by a suitable normalization of parameters, these properties can be represented in a convenient universal form in either the complex dielectric coefficient plane or the complex propagation constant plane. As an illustration, the propagation characteristics of high-temperature air represented in the propagation plane are shown for an impressed frequency of 6 kmc.

UNIVERSAL REPRESENTATION OF ELECTROMAGNETIC PARAMETERS

It is instructive to write the relationships determining the complex dielectric coefficient of a plasma and propagation constant of an incident plane electromagnetic wave on a plasma in normalized form, and hence demonstrate their behavior in universal coordinates. Thus, we define the following parameters.

- 1) normalizing with respect to frequency:

$$S = \nu/\omega = \text{normalized scattering frequency,} \\ N = (\omega_p/\omega)^2 = \text{normalized electron density,}$$

where

ω is the radian frequency of the electromagnetic wave,
 ν is the effective collision frequency of the electrons,

ω_p is the parameter ($ne^2/\epsilon_0 m$) which has the dimensions of seconds⁻¹ and is called the plasma frequency, e and m being the electronic charge and mass, respectively, n the electron density and ϵ_0 the permittivity of free space;

- 2) normalizing with respect to plasma frequency (*i.e.*, $n^{1/2}$):

$$C = \nu/\omega_p = \text{normalized collision parameter,} \\ F = \omega/\omega_p = \text{normalized RF frequency.}$$

These relationships permit the mapping of loci of constant scattering frequency S , constant electron density N , constant collision parameter C , or constant RF frequency F on the complex dielectric coefficient plane and on the complex propagation constant plane. S and N are the useful parameters to consider in a diagnostic measurement, with a given frequency and varying plasma, while C and F are useful when the frequency behavior of a given plasma is of interest.

Dielectric Plane

Using the relationship

$$K = \left\{ 1 - (\omega_p/\omega)^2 \frac{1}{1 + (\nu/\omega)^2} \right\} \\ - j \left\{ (\omega_p/\omega)^2 \frac{(\nu/\omega)}{1 + (\nu/\omega)^2} \right\} \\ = Kr + jKi \quad (1)$$

for the complex dielectric coefficient, it is easily seen that the normalized scattering term is given by

$$S = Ki/(Kr - 1), \quad (2a)$$

or

$$(Kr - 1)S + Ki = 0, \quad (2b)$$

which is a family of straight lines in the complex dielectric (Kr vs Ki) plane, with slope ($-S$) and Kr -intercept of 1.

Similarly, the normalized electron density is

$$N = 1 - Kr + Ki^2/(1 - Kr), \quad (3a)$$

or

$$(Kr - (1 - N/2))^2 + Ki^2 = (N/2)^2, \quad (3b)$$

which in the complex plane is a family of circles of radius $N/2$ and center [$Kr = (1 - N/2)$, $Ki = 0$].

Representation in terms of the normalized collision parameter is slightly more difficult in that

$$C^2 = S^2/N, \quad (4a)$$

† Research Laboratories, RCA Victor Co., Ltd., Montreal, Can.

or in terms of the real and imaginary part of the dielectric coefficient

$$(1 - Kr)^3 + Ki^2(1 - Kr) - Ki^2/C^2 = 0 \quad (4b)$$

or

$$Ki = (1 - Kr) \left(\frac{1}{C^2(1 - Kr)} - 1 \right)^{-1/2} \quad (4c)$$

with a pole at $Kr = 1 - 1/C^2$.

The normalized RF frequency loci are again circles in the complex dielectric coefficient plane, since

$$F = 1/\sqrt{N}. \quad (5)$$

The radius of the F -circles is $1/1$ and their centers are located at $(Kr = (1 - 1/(2F^2)), Ki = 0)$.

Propagation Plane

Representation of the normalized parameters in the complex propagation constant plane can be derived from a conformal transformation of values in the dielectric plane, or by directly solving for the normalized parameters in terms of the attenuation and phase constants, since the propagation constant is defined by

$$\gamma = -jkK^{1/2} = \alpha + j\beta, \quad (6a)$$

where

$$\frac{\alpha}{k} = \sqrt{\frac{|K| - Kr}{2}} = A \quad (6b)$$

$$\frac{\beta}{k} = \sqrt{\frac{|K| + Kr}{2}} = B \quad (6c)$$

(α is called the attenuation constant, β the phase constant of the plasma, and k is the wave number $= 2\pi$ /wavelength). Since the maximum value of the real part of the dielectric constant is unity, the upper limit in the propagation plane is given by the line $Kr = 1$, which maps into the hyperbola: $B^2 - A^2 = 1$.

The normalized scattering frequency loci become rectangular hyperbolas rotated through an angle of $\frac{1}{2} \tan^{-1}(1/S)$ in the complex propagation constant plane, namely:

$$S = \frac{2AB}{1 - (B^2 - A^2)}, \quad (7a)$$

or

$$B^2 - A^2 + 2AB/S = 1. \quad (7b)$$

The normalized electron density families become quartic curves, since

$$N = 1 - A^2 - B^2 + 4A^2B^2/(1 + A^2 - B^2), \quad (8a)$$

$$(A^2 + B^2)^2 - (2 - N)(B^2 - A^2) + (1 - N) = 0. \quad (8b)$$

The normalized collision parameter, C , can be expressed in terms of A and B by the use of (7) and (8a) in (4a). Similarly the values of normalized frequency F in the propagation plane is obtained by substituting (8a) into (5).

It will be observed that the normalized parameters map more readily into the complex dielectric coefficient plane than into the propagation constant plane. However, in measurement or diagnostics it is the attenuation and phase constants which are actually determined. Hence, the added difficulty in plotting the normalized contours in the propagation constant plane is generally justified.

A second reason for representing the plasma parameters in the propagation plane is that the reflecting or transmitting properties of a plasma boundary are very conveniently mapped in the propagation plane. Thus, if Z is the impedance of the plasma and Z_0 the impedance of free space, the fraction of the field of a normally incident electromagnetic wave reflected at a plasma free-space interface is given by the reflection coefficient R , where

$$R = \frac{Z - Z_0}{Z + Z_0} \quad (9a)$$

$$= \frac{(1 - B) + jA}{(1 + B) - jA}. \quad (9b)$$

The magnitude of the reflection coefficient is given by

$$|R| = \sqrt{\frac{1 + A^2 + B^2 - 2B}{1 + A^2 + B^2 + 2B}} \quad (10a)$$

$$= \sqrt{\frac{1 - x}{1 + x}}, \quad (10b)$$

where

$$x = 2B/(1 + A^2 + B^2). \quad (10c)$$

Similarly, the magnitude of the transmission coefficient is

$$|T| = \sqrt{1 - |R|^2} = \sqrt{\frac{2x}{1 + x}}. \quad (10d)$$

For any magnitude of reflection coefficient, the relationship between the attenuation A and the phase B is uniquely determined by (10c), which can be rewritten as

$$A^2 + (B - 1/x)^2 = \left(\frac{1}{x^2} - 1\right), \quad (10e)$$

which is a family of circles with center ($A=0$, $B=1/x$) and radius

$$\left(\frac{1-x^2}{x^2}\right)^{1/2}.$$

If, in a particular application, the maximum allowable attenuation and reflection coefficient are known, then the "operating region" in the propagation plane is defined by the area enclosed by the B -axis and the lines $B^2 - A^2 = 1$, $R = R_{\max}$ and $\alpha/k = \alpha_{\max}/k$. Similar plots and operating regions may be determined for the transmission ($T_{\min} = \sqrt{1 - |R_{\max}|^2}$) properties of a plasma.

The propagation characteristics of high-temperature air¹ are represented in the propagation plane for an impressed frequency of 6 kmc in Fig. 1. Contours of constant temperature, constant density, and constant reflection coefficient are shown. At low temperatures, the plasma behaves nearly like free space at all densities. As the temperature increases, the influence of the ambient density becomes more apparent as the plasma becomes more lossy. At high temperatures, the plasma is a good conducting medium and the effect of density variation again becomes secondary. In this representation, an operating region for propagation of electromagnetic energy can be determined, provided the tolerable attenuation and reflection coefficients are specified.

¹ For methods of determining the properties of high-temperature air and further extensive calculations, see M. P. Bachynski, I. P. Shkarofsky, and T. W. Johnston, "Plasma Physics of Shock Fronts," RCA Res. Labs., Montreal, Can., Res. Rept. No. 7-801-3; June, 1959.

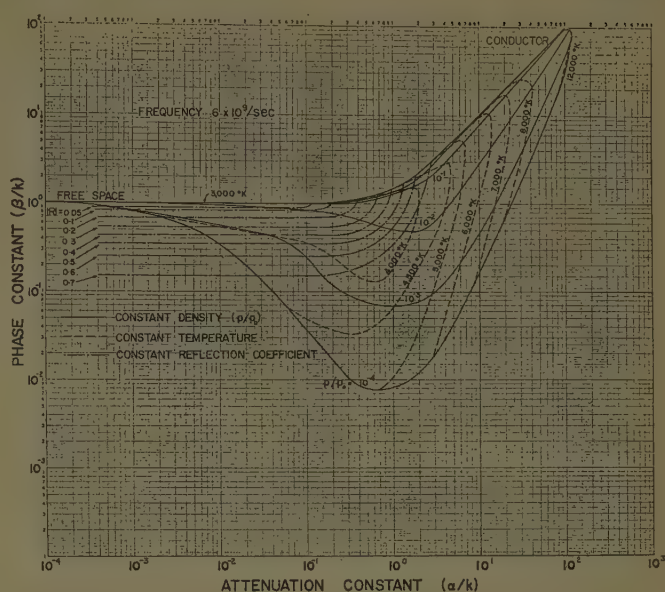


Fig. 1—Electromagnetic properties of high-temperature air at a frequency of 6×10^9 cps, showing the variation of the attenuation and phase constants at constant temperature, constant density, and constant reflection coefficient.

Similar plots can be made for the plasma sheath surrounding a vehicle traversing the atmosphere, since the temperature can be related to the vehicle velocity and altitude to density.

ACKNOWLEDGMENT

The authors are indebted to the Aerophysics Wing of the Canadian Armament and Research Development Establishment for financial support.

The Propagation of Electromagnetic Waves in Ionized Gases*

Part I—Introductory Theory

F. H. NORTHOVER†

Summary—Recent studies lend support to the theory that whistling atmospherics are caused by lightning flashes, the electromagnetic energy radiated by these being guided along discrete columnar ionic irregularities which follow approximately the lines of force of the earth's magnetic field. In Part I, the theoretical problems that arise are set forth and a general wave theory is developed which is first applied to the problem of propagation through homogeneous compound streaming media. In Part II, the simplest case of "standard type" propagation along stationary columns is carefully examined, both for columns with a central ionic surplus and for columns with a central ionic deficiency. Although both types of columns can guide electromagnetic energy when sufficiently well developed, it appears that the former type is a much more likely mechanism for the whistler propagation than the latter. In Part III, an attempt is made to show how a systematic theory might be developed for the case of axially moving columns. It is hoped in subsequent papers to further extend this theory and also to deal with propagation of a more general type of disturbances along the columns.

I. INTRODUCTION

IT is now fairly well established¹ that whistling atmospherics are caused by lightning flashes, the electromagnetic energy radiated by these being guided fairly well along the geomagnetic field lines through the ionosphere.

This guidance alone does not seem sufficient to explain the high intensities sometimes observed in multiple or very long distance single paths, but a waveguide effect might exist if the ionization were suitably bunched in columns along the field lines, and such an effect might explain the observed intensities. Further, this bunching could provide discrete preferred paths of propagation, each with its own travel time, which would account for the discrete dispersion traces frequently observed in the frequency against time of transit records of whistling atmospherics.

There is reason to believe that such columns might exist, because electrons in a constant magnetic field have a natural tendency to execute helical paths whose axes are parallel to the magnetic field. Progression parallel to the field is at a constant rate, and the number of circuits per second about the helix axis is equal to the gyro-frequency. The suggestion is, that in the upper reaches of the ionosphere, where collisions are relatively infrequent, this tendency causes inhomogeneities of charged

particles to diffuse (in approximately helical paths) along some of the lines of force of the earth's magnetic field, ultimately forming more or less complete columns of ionization. It therefore seems worthwhile to investigate the propagation of electromagnetic waves in the presence of such columns.

In the applications to columns of the general analysis which will be developed in this paper, the curvature of the axis of the column will be neglected; we shall suppose that the column extends to infinity in both directions and shall neglect possible variations of ionic density in axial directions. As far as our wave treatment is concerned, this is equivalent to supposing that the axial curvature is very small compared with the reciprocal of the wavelength in the medium and that the percentage variations of ionic density along the axis are extremely small over many wavelengths.

It seems likely that the type of column to be considered consists of a compound cylindrically distributed ionic distribution consisting of a neutral ionized gas streaming through a neutral background of stationary ionized gas. If N is the ionic density of the moving gas and \bar{N} this quantity for the stationary gas, U the stream velocity of the moving gas and C the velocity of light, then the case of practical interest in regard to whistlers is that of N/\bar{N} small, U/C small. If U/C is sufficiently small we have effectively a stationary columnar distribution, and, in view of the extreme difficulty of the whole problem we shall make the first application of the general theory concerning columns to this case.

A. Plane-Wave Incident Standard-Type Disturbances

Considering, for the moment, the modes of propagation down a stationary homogeneous cylindrical column (N constant through the column, N equal to a different constant outside), the disturbance within the column, as the radius of the cross section tends to infinity, must approximate those possible in a homogeneous medium of infinite extent. The simplest of these latter types of disturbance is the plane-wave mode with direction of travel parallel to the original column's axis. The disturbance, propagated down the actual column, which would tend to this simple type were the column radius to tend to infinity, will be called the "plane wave incident" or "standard" type disturbance. It will be taken as the standard type of disturbance in the present theory.

The standard disturbance propagated down a non-homogeneous axially symmetric ionized column would

* This research was supported by contributions from the Radio Physics Laboratory of the Defence Research Telecommunications Establishment, Defence Research Board, Ottawa, Can., under PCC D48-95-11-01, and the Department of Mathematics, Carleton University, Ottawa, Can.

† Dept. of Mathematics, Carleton University, Ottawa, Can.

¹ See Bibliography at the conclusion of Part III.

be defined as the one which would reduce to the above basic axially traveling plane wave, when $dN/dr \rightarrow 0$ for every r , r being the distance from the axis of a general point in the distribution.

In the more complicated case of the compound distribution, it appears that plane-wave propagation is generally possible when both constituent distributions are homogeneous. Standard-type disturbances can therefore be defined in a similar fashion. The theory would thus seem to fall naturally into the following divisions.

1) *Propagation Down a Simple Columnar Distribution*: Standard and nonstandard type propagation in nonhomogeneous and homogeneous columns.

2) *Propagation Down a Compound Columnar Distribution*: Standard and nonstandard propagation down nonhomogeneous and homogeneous columns.

After general equations governing wave propagation of electromagnetic disturbances through ionized gases have been developed, the writer will endeavor to deal with some of these problems in a series of papers.

The original magneto-ionic theory as developed by Appleton and Hartree is not sufficiently general for our requirements. We shall therefore employ magneto-hydrodynamical methods. Since the simple type of distribution is a special case ($U=0$) of the compound distribution, we shall begin by developing the general equations for the latter, although the actual application will be to the simple-type stationary distributions first. This will obviate the necessity of developing a general-type theory twice.

II. GENERAL THEORY OF WAVES IN CONDUCTING GASES

The instantaneous current vector \mathcal{J} at any point due to the passage of the wave consists of four parts; namely, the contributions due to displacements of the positive and negative ions in the steadily streaming gas and the contributions due to displacements of the two kinds of ions in the "background" gas through which the former gas is flowing. The value of \mathcal{J} thus calculated can then be used in Maxwell's equations of the field to obtain a wave equation for the electric vector \mathbf{E} .

A. Detailed Consideration of the Moving Gas

Let ρ be the gas density, $Td\tau$ the resultant force upon a volume element $d\tau$, \mathbf{W} the velocity of this element and Π the gas pressure. Then the hydrodynamical equations are²

$$\frac{D\mathbf{W}}{Dt} = \mathbf{T} - \frac{1}{\rho} \text{grad } \Pi \quad (1)$$

where D/Dt denotes "differentiation following the fluid"; i.e.,

$$D/Dt = \partial/\partial t + \mathbf{W} \cdot \text{grad}$$

and

$$\text{div}(\rho \mathbf{W}) = -\partial \rho / \partial t. \quad (2)$$

Since we are supposing that the gas, although ionized, is electrically neutral, a volume element $d\tau$ (taken to contain a large number $2Nd\tau$ of particles) contains $Nd\tau$ electrons and $Nd\tau$ protons.

The vibration of the electrons and protons is caused by the passing electromagnetic wave. The mass of a proton is considerably greater than that of an electron and so, for frequencies considerably greater than the proton gyro-frequency and not too near the proton plasma frequency, the wave is carried almost entirely by the electrons, the positive particles remaining almost stationary in comparison. We shall therefore concentrate mainly upon the propagation in the electron gas, although, for completeness, the equations of the propagation in the complete medium will be noted. This would be of use in discussion of the frequencies excepted above.

We proceed to the derivation of the equations of the theory. The force per unit mass acting upon an element $d\tau$ either of electron or proton gas is compounded as follows:

1) *The Body Force per Unit Mass, \mathbf{F} , Say, Due to Nonelectromagnetic Causes*: In terrestrial problems, this may be taken as the local value of \mathbf{g} . This, of course, neglects the effect of the internal gravitational attraction of the gas on itself: this may become important in cosmogonical problems but it is not important here.³

2) *Collisional Effects*: As is now well known, this can be represented by a friction-type force on any given particle proportional to its velocity through the gas. If m is the mass of a negative ion, ν the collision frequency, \mathbf{V}_i its velocity and \mathbf{U} the velocity of the background material, then this effect can be allowed for by a force $-m\nu(\mathbf{V}_i - \mathbf{U})$. Summing over the particles in the volume element, the contribution of this type of force to the element $d\tau$ of electron gas is $-m\nu(\mathbf{V} - \mathbf{U})Nd\tau$, where \mathbf{V} is the average velocity of all the particles in the volume element. From considerations of momentum, it is easy to see that \mathbf{V} is also the effective (hydrodynamical) velocity of the gas element $d\tau$.

Similarly, denoting by an accent quantities pertaining to the proton gas, the force on a volume element $d\tau$ of it due to this cause will be

$$-m'\nu'(\hat{\mathbf{V}} - \mathbf{U})Nd\tau.$$

We can remark here the connection between the hydrodynamical velocity \mathbf{W} of the element $d\tau$ of the complete (neutral) gas and the corresponding velocities \mathbf{V} , $\hat{\mathbf{V}}$ of the constituent negatively and positively charged

² We assume that the gases are sufficiently rarified for the effects of viscosity to be unimportant; although, of course, the analysis could be carried through with the viscosity term present.

³ A note on the calculation of this effect will be given later.

gases. From a consideration of the total momentum of the element $d\tau$ of the neutral gas, we have

$$m\mathbf{V} + \hat{m}\hat{\mathbf{V}} = (m + \hat{m})\mathbf{W}. \quad (3)$$

3) *Force Produced on the Element by the Magnetic Field:* Considering first the negatively charged gas only, the force exerted by the magnetic field \mathbf{B} upon a constituent particle is (neglecting relativistic effects) $-(e/C)(\mathbf{V}_i \times \mathbf{B}_i)$, \mathbf{B}_i being the magnetic induction vector at the i th particle. Summing this over the particles in the element $d\tau$, the total force produced on the element $d\tau$ of electron gas through this cause, is $-(e/C)Nd\tau(\mathbf{V} \times \mathbf{B})$, \mathbf{B} denoting the magnetic field at the element.

The corresponding force for the positively charged gas will, of course, be

$$+(e/C)Nd\tau(\mathbf{V} \times \mathbf{B}).$$

4) *Force due to the Electric Field \mathbf{E} :* This is obviously $-(Ned\tau)\mathbf{E}$ for the element of electron gas and $+(Ned\tau)\mathbf{E}$ for the proton gas.

The hydrodynamical equation of motion for the electron gas may now be stated. It is

$$\begin{aligned} \frac{D\mathbf{V}}{Dt} = & -\frac{e}{m}\mathbf{E} - \nu(\mathbf{V} - \mathbf{U}) - \frac{e}{mC}\mathbf{V} \wedge \mathbf{B} \\ & - \frac{1}{mN} \text{grad } p + \mathbf{F} \end{aligned} \quad (4)$$

where p here is the electron pressure (and so the force it exerts on $d\tau$ is $-d\tau \text{ grad } p$). Likewise, the corresponding equation for the proton gas is

$$\begin{aligned} \frac{D\hat{\mathbf{V}}}{Dt} = & \frac{e}{\hat{m}}\mathbf{E} - \hat{\nu}(\hat{\mathbf{V}} - \mathbf{U}) + \frac{e}{\hat{m}C}\hat{\mathbf{V}} \wedge \mathbf{B} \\ & - \frac{1}{\hat{m}N} \text{grad } \hat{p} + \mathbf{F} \end{aligned} \quad (5)$$

where \hat{p} is the proton gas pressure.

Since the electron gas density is mN , the hydrodynamical equation of continuity is

$$\text{div}(\mathbf{NV}) = -\partial N/\partial t. \quad (6)$$

But the current vector \mathbf{I} of the electron gas is $-eN\mathbf{V}$ and the electric charge in the volume $d\tau$ for this gas is $-eNd\tau$. Hence, the electrical equation of continuity is the same as the hydrodynamical one and we shall therefore refer to (6) without ambiguity as "the equation of continuity" (for the negatively charged gas). Similarly, the equation of continuity of the proton gas will be

$$\text{div}(\mathbf{N}\hat{\mathbf{V}}) = -\partial N/\partial t. \quad (7)$$

Using these equations and (3) we obtain, as we should, the equation of continuity for the complete (neutral) gas, *viz.*,

$$\text{div}(\mathbf{NW}) = -\partial N/\partial t. \quad (8)$$

In the passage of a small wave disturbance through the steadily streaming neutral material $\mathbf{V} - \mathbf{U}$, $\hat{\mathbf{V}} - \mathbf{U}$ will be small and so, neglecting second-order small quantities, we can take

$$D/Dt \doteq \partial/\partial t + (\mathbf{U} \cdot \text{grad}). \quad (9)$$

Eqs. (4) and (5), when used in conjunction with (3), yield an expression for DW/dt and then, eliminating this quantity by (1) we can prove

$$\Pi = p + \hat{p} + \frac{Ne}{C} \int (\mathbf{V} - \hat{\mathbf{V}}) \times \mathbf{B} \cdot d\mathbf{r} \quad (10)$$

so that Dalton's law of partial pressures applies in the waves to a first approximation.

We now have to apply this theory to the passage of such a small disturbance, and the equations governing the passage of the electromagnetic wave through the medium will be obtained by a perturbation process; *i.e.*, by comparing (4) and (5) with the corresponding equations for the steady state. Since we are neglecting the effect of the internal gravitational pulls of the gas on itself, in this perturbation, $\delta\mathbf{F} = 0$. Also, writing

$$\mathbf{V} = \mathbf{U} + \mathbf{v} \quad (11)$$

$$\mathbf{B} = \mathbf{H}_0 + \mathbf{H} \quad (12)$$

(where we are taking the permeability as unity and \mathbf{H} as the magnetic field due to the wave), \mathbf{H}_0 is the external (applied) magnetic field (effectively the earth's field) and \mathbf{v} , \mathbf{H} are small quantities. Taking suffix zero to refer to steady-state values and retaining only first order small quantities, we have (for the negatively charged gas)

$$\begin{aligned} \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{U} \cdot \text{grad})\mathbf{v} \\ = & -\frac{e}{m}\mathbf{E} - \nu\mathbf{v} - \frac{e}{mC}\mathbf{v} \wedge \mathbf{H}_0 - \frac{e}{mC}\mathbf{U} \wedge \mathbf{H} \\ & + \frac{\delta N}{mN_0^2} \text{grad } p - \frac{1}{mN_0} \text{grad}(\delta p) \end{aligned} \quad (13)$$

and

$$\partial(\delta N)/\partial t + \text{div}\{(\delta N)\mathbf{U}\} + \text{div}(N_0\mathbf{v}) = 0. \quad (14)$$

The current vector \mathbf{I} in the electron gas due to the wave is

$$\begin{aligned} \mathbf{I} = & -(eN\mathbf{V} - eN_0\mathbf{U}) \\ = & -e\{(\delta N)\mathbf{U} + N_0\mathbf{v}\} \end{aligned} \quad (15)$$

so that

$$\text{div } \mathbf{I} = e\partial(\delta N)/\partial t. \quad (16)$$

Let

$$\beta^2 = (dp/d\rho)_0 \quad (17)$$

p being taken as a function of ρ (it would probably be reasonable to take this as the adiabatic relation, as

in sound-wave theory). Then, neglecting second- and higher-order small quantities, we have

$$\begin{aligned}\delta p &= \beta^2 \delta \rho \\ &= \beta^2 m \delta N\end{aligned}\quad (18)$$

and the first of the above perturbed equations becomes

$$\begin{aligned}\dot{\mathbf{v}} + (\mathbf{U} \cdot \text{grad}) \mathbf{v} \\ = -\frac{e}{m} \mathbf{E} - \nu \mathbf{v} - \frac{e}{mC} \mathbf{v} \wedge \mathbf{H}_0 - \frac{e}{mC} \mathbf{U} \wedge \mathbf{H} \\ + \frac{\delta N}{mN_0^2} \text{grad } p_0 - \frac{\beta^2}{N_0} \text{grad } (\delta N).\end{aligned}\quad (19)$$

The term $\text{grad } p_0$ depends upon the body forces \mathbf{F} which hold the material in the steady state. Writing into (4) steady-state values we have

$$0 = -\frac{e}{mC} \mathbf{U} \wedge \mathbf{H}_0 - \frac{1}{mN_0} \text{grad } p_0 + \mathbf{F}.\quad (20)$$

Hence (19) is, effectively, (writing $\mathbf{F} \doteq \mathbf{g}$)

$$\begin{aligned}\dot{\mathbf{v}} + (\mathbf{U} \cdot \text{grad}) \mathbf{v} \\ = -\frac{e}{m} \mathbf{E} - \nu \mathbf{v} - \frac{e}{mC} \mathbf{v} \wedge \mathbf{H}_0 - \frac{e}{mC} \mathbf{U} \wedge \mathbf{H} \\ + \frac{\delta N}{N_0} \left(\mathbf{g} - \frac{e}{mC} \mathbf{U} \wedge \mathbf{H}_0 \right) - \frac{\beta^2}{N_0} \text{grad } (\delta N).\end{aligned}\quad (21)$$

B. Specialization for Sound Waves Only

Writing the electrical terms and \mathbf{U} equal to zero and using the equation of continuity, we may verify that, in this special case,

$$\begin{aligned}\partial^2(\delta N)/\partial t^2 + \nu \partial(\delta N)/\partial t + \mathbf{g} \cdot \text{grad } (\delta N) \\ = \beta^2 \text{div grad } (\delta N).\end{aligned}\quad (22)$$

If the frictional and gravitational terms are neglected, this becomes the usual scalar sound-wave equation. It may readily be verified that this approximation may be made if the frequency makes $\omega \gg g/\beta$ and $\omega \gg \nu$. We shall make this approximation in our application of the general theory which follows.

C. Determination of the Current Vectors

Concentrating first on the electron gas and its current vector \mathbf{I} , after a certain amount of reduction these equations show that

$$\begin{aligned}\left(\frac{\partial}{\partial t} + \mathbf{U} \cdot \text{grad} + \nu \right) \{ \dot{\mathbf{I}} + (\text{div } \mathbf{I}) \mathbf{U} \} \\ = \frac{N_0 e^2}{m} \dot{\mathbf{E}} - \frac{e}{mC} \dot{\mathbf{I}} \wedge \mathbf{H}_0 + \frac{N_0 e^2}{mC} \mathbf{U} \wedge \dot{\mathbf{H}} - (\text{div } \mathbf{I}) \mathbf{g} \\ + \beta^2 \text{grad div } \mathbf{I}.\end{aligned}\quad (23)$$

Writing

$$\omega_0^2 = 4\pi N_0 e^2/m, \quad \omega_H = |\mathbf{H}_0|/mC \quad (24)$$

\mathbf{n} is equal to a unit vector along \mathbf{H}_0 . Taking Oz along the direction of \mathbf{U} , and writing all vectors proportional to $\exp(i\omega/C\sqrt{\kappa}z - i\omega t)$,⁴ we have

$$\begin{aligned}\left(-i\omega + \frac{i\omega}{C} \sqrt{\kappa} U + \nu \right) \{ -i\omega \mathbf{I} + (\text{div } \mathbf{I}) \mathbf{U} \} \\ = -\frac{1}{4\pi} i\omega \omega_0^2 \left(\mathbf{E} + \frac{1}{C} \mathbf{U} \wedge \mathbf{H} \right) + i\omega \omega_H \mathbf{I} \wedge \mathbf{n} \\ - (\text{div } \mathbf{I}) \mathbf{g} + \beta^2 \text{grad div } \mathbf{I}.\end{aligned}\quad (25)$$

The equation for $\hat{\mathbf{I}}$, the current vector in the positively charged gas, will be obtained from this by replacing \mathbf{I} by $\hat{\mathbf{I}}$, β by $\hat{\beta}$ and changing ω_H into $-\hat{\omega}_H$. To obtain the equations for the current vectors in the stationary gas, e.g., $\bar{\mathbf{I}}$ for the negatively charged part and $\bar{\hat{\mathbf{I}}}$ for the positively charged part, we can write $\mathbf{U} = 0$ in the above and make the necessary alterations. We thus have

$$\begin{aligned}(i\omega - \nu) i\omega \bar{\mathbf{I}} = -\frac{1}{4\pi} i\omega \bar{\omega}_0^2 \mathbf{E} + i\omega \omega_H \bar{\mathbf{I}} \wedge \mathbf{n} \\ - (\text{div } \bar{\mathbf{I}}) \mathbf{g} + \bar{\beta}^2 \text{grad div } \bar{\mathbf{I}}\end{aligned}\quad (26)$$

$$\begin{aligned}(i\omega - \nu) i\omega \bar{\hat{\mathbf{I}}} = -\frac{1}{4\pi} i\omega \bar{\omega}_0^2 \mathbf{E} - i\omega \hat{\omega}_H \bar{\hat{\mathbf{I}}} \wedge \mathbf{n} \\ - (\text{div } \bar{\hat{\mathbf{I}}}) \mathbf{g} + \bar{\hat{\beta}}^2 \text{grad div } \bar{\hat{\mathbf{I}}}.\end{aligned}\quad (27)$$

If we neglect the pressural and body force terms in (26), we obtain an equation which, when solved for $\bar{\mathbf{I}}$ and used in (32) with \mathbf{g} therein taken approximately equal to $\bar{\mathbf{I}}$, embodies the results of the classical magneto-ionic theory for waves through stationary media.

1) *Note on the Correction for Internal Gravitation:* If this is allowed for, then the perturbation $\delta \mathbf{F}$ is no longer zero but satisfies the relation (if stationary and streaming gases exist together)

$$\text{div } (\delta \mathbf{F}) = -4\pi \gamma (m + \hat{m}) (\delta N + \delta \bar{N}) \quad (28)$$

where γ is the gravitation constant. It is now possible to deduce the rather elegant result that the required correction will be made by adding to the right sides of (25)–(27) the term $-4\pi \gamma (\rho_0 + \bar{\rho}_0) (\mathbf{I} + \bar{\mathbf{I}})$ where ρ_0 is the undisturbed mass density of the streaming gas and $\bar{\rho}_0$ is the undisturbed mass density of the stationary gas.

D. Derivation of the Equation Connecting \mathbf{E} with \mathbf{g}

Maxwell's field equations are

$$\text{curl } \mathbf{H} = \frac{4\pi}{C} \mathbf{g} + \frac{1}{C} \dot{\mathbf{D}} \quad (29)$$

$$\text{curl } \mathbf{E} = -\frac{1}{C} \dot{\mathbf{B}} \quad (30)$$

together with

$$\text{div } \mathbf{D} = 4\pi \rho, \quad \text{div } \mathbf{B} = 0. \quad (31)$$

⁴ This is appropriate for propagation along columns.

From these equations, taking (as usual) the magnetic permeability as unity and also $\mathbf{D}=\mathbf{E}$ in the conducting gas (since the electrons, etc., are taken as vibrating in vacuo) and, writing, as above, all vectors proportional to $\exp(-i\omega t)$, we obtain

$$\text{curl curl } \mathbf{E} - \frac{\omega^2}{C^2} \mathbf{E} = \frac{4\pi i\omega}{C^2} \mathcal{G} \quad (32)$$

and connection with the preceding equations is made through the equation for the total current vector

$$\mathcal{G} = \mathbf{I} + \hat{\mathbf{I}} + \bar{\mathbf{I}} + \bar{\bar{\mathbf{I}}}. \quad (33)$$

The problem is now theoretically solved, for if (25)–(27) could be solved for the four constituent current vectors, we should, by use of (35) in (32) and by use of the relation

$$\mathbf{H} = \frac{iC}{\omega} \text{curl } \mathbf{E} \quad (34)$$

to eliminate \mathbf{H} , obtain an equation in the electric field \mathbf{E} alone.

The practical difficulties are obviously immense and we shall therefore proceed tentatively. We shall begin by investigating the propagation of plane waves along the Oz axis (direction of \mathbf{U}) in the special case when \mathbf{H}_0 is also along this direction. This is the case which is of importance in whistling atmospherics. We shall then go on to examine propagation along columns, both simple stationary ones and the compound streaming types. In this connection, it is very interesting to remark that, for a stationary column, when the effect of the positively charged particles are neglected, it is possible to solve (26) for $\bar{\mathbf{I}}$. For in that case, $\text{div } \bar{\mathbf{I}}$ can be expressed directly in terms of $\text{div } \mathbf{E}$ from (32), by taking \mathcal{G} as effectively $\bar{\mathbf{I}}$.

III. THE PROPAGATION OF PLANE WAVES⁵

We suppose that the frequency is considerably greater than the proton gyro-frequency and is also not near the proton plasma frequency. In that case, we are neglecting $\hat{\mathbf{I}}$ and $\bar{\bar{\mathbf{I}}}$ in comparison with \mathbf{I} and $\bar{\mathbf{I}}$. Expressing our vector equations in components, with all vectors varying as $\exp\{(i\omega/C)\sqrt{\kappa}z - i\omega t\}$, we find that two different types of waves which are quite independent of each other can be propagated. These are the following.

A. The Sound-Wave Type ($E_x=0$, $E_y=0$, $E_z \neq 0$)

The propagation constant κ satisfies the equation

$$\frac{\omega_0^2}{\left(1 - \frac{U}{C}\sqrt{\kappa}\right)^2 - \frac{\beta^2}{C^2}\kappa} + \frac{\bar{\omega}_0^2}{1 - \frac{\bar{\beta}^2}{C^2}\kappa} = \omega^2. \quad (35)$$

We call this the "sound-wave" mode because, if we have a single stationary medium ($N=0$, $\omega_0=0$), then de-

electrify it ($e \rightarrow 0$, $\bar{\omega}_0 \rightarrow 0$) and neglect the internal gravitational term in γ , the phase velocity $C/\sqrt{\kappa} \rightarrow \bar{\beta}$, as it should. We note that this mode is independent of the magnetic field, and, furthermore, that, for the stationary medium case, it is heavily attenuated for frequencies less than the medium plasma frequency.

B. The Electromagnetic Type ($E_z=0$)

The propagation constant satisfies the equation

$$\kappa - 1 - \frac{\bar{\omega}_0^2}{\omega(\omega_H - \omega)} = \frac{\left(1 - \frac{U}{C}\sqrt{\kappa}\right)\omega_0^2/\omega^2}{\frac{\omega_H - \omega}{\omega} + \frac{U}{C}\sqrt{\kappa}}. \quad (36)$$

Having whistlers in mind, we have taken the solution that reduces to

$$\kappa = 1 + \frac{\bar{\omega}_0^2 + \bar{\omega}_0^2}{\omega(\omega_H - \omega)} \quad \text{if } U \rightarrow 0.$$

From this equation, it appears that, if the direction of propagation is against the stream velocity ($U < 0$), there is always one negative solution for $\sqrt{\kappa}$ (backward traveling wave) and either two positive solutions (forward traveling waves) or two conjugate complex solutions. If $U > 0$ (propagation *with* the stream) there is always one forward traveling wave ($\sqrt{\kappa} > 0$) whereas there may or may not exist two backward traveling waves. This discussion is confined to the question of whether there are any circumstances under which forward traveling waves could go out of existence; therefore limiting it to the case of propagation *against* the stream. The two positive roots for $\sqrt{\kappa}$ from (36) disappear if

$$4(1 - b\epsilon_1)^3(a^2 + b)^2\alpha^4 - b^3(a^2 + b)\alpha^2\{8 + 4\epsilon_1(5b + 9) - \epsilon_1^2(b^2 - 18b - 27)\} + 4b^6(1 + \epsilon_1) < 0 \quad (37a)$$

where

$$\begin{aligned} \alpha &= U/C, & b &= (\omega_H - \omega)/\omega, & a^2 &= \bar{\omega}_0^2/\omega^2 \\ \epsilon &= \omega_0^2/\bar{\omega}_0^2 (= N_0/\bar{N}_0) \\ \epsilon_1 &= \left(1 + \frac{b}{a^2}\right)^{-1} \epsilon. \end{aligned} \quad (37b)$$

Writing

$$\kappa_0 = 1 + \frac{a^2}{b} \quad (38)$$

(the value of κ when the moving medium is away) so that

$$\epsilon_1 = (1 - \kappa_0^{-1})\epsilon \quad (39)$$

the above condition for the two positive roots for $\sqrt{\kappa}$ to disappear in the case $U < 0$ is equivalent to

⁵ and the term in g in this theory are neglected.

$$\left\{ \frac{8 + 4\epsilon_1(5b + 9) - (b^2 - 18b - 27)\epsilon_1^2 - [(b + 9)\epsilon_1 + 8]^{3/2}\sqrt{\epsilon_1(b + 1)}}{8(1 - b\epsilon_1)^3} \right\}^{1/2} < \left| \frac{\alpha\sqrt{\kappa}}{b} \right|$$

$$< \left\{ \frac{8 + 4\epsilon_1(5b + 9) - (b^2 - 18b - 27)\epsilon_1^2 + [(b + 9)\epsilon_1 + 8]^{3/2}\sqrt{\epsilon_1(b + 1)}}{8(1 - b\epsilon_1)^3} \right\}^{1/2} \quad [b\epsilon_1 < 1], \quad (40)$$

$$\left| \frac{\alpha\sqrt{\kappa_0}}{b} \right| > \frac{2}{3\sqrt{3}} \quad [b\epsilon_1 = 1] \quad (41)$$

$$\left| \frac{\alpha\sqrt{\kappa_0}}{b} \right| > \left\{ \frac{[(b + 9)\epsilon_1 + 8]^{3/2}\sqrt{\epsilon_1(b + 1)} - \{8 + 4\epsilon_1(5b + 9) - (b^2 - 18b - 27)\epsilon_1^2\}}{8(b\epsilon_1 - 1)^3} \right\}^{1/2} \quad [b\epsilon_1 > 1]. \quad (42)$$

The case of most practical interest occurs when $\omega < \omega_H$, $\epsilon \ll 1$, $a^2 \gg b$. Some simplification exists in the above showing that there is then a narrow band of values for $|\alpha|$, for which the two positive values of $\sqrt{\kappa}$ representing the two forward traveling waves become complex, unless, possibly, ω/ω_H becomes so small that $\epsilon\omega_H/\omega$ is no longer small. For example, if $\epsilon\omega_H/\omega$ becomes large, the condition becomes $|\alpha| > \epsilon^{-1/2}(\omega_H/2\tilde{\omega}_0)$, which probably could not be satisfied in practice.

The question of the physical meaning of these complex roots now arises, because, being complex numbers, one represents a disturbance which rapidly grows as z increases until the point is reached when this first-order theory fails. The writer has had considerable discussion of this with colleagues and there seems to be an element of uncertainty as to whether this "growing wave" can really exist. The writer thinks not, for the reasons which follow.

If a physical disturbance of the ionized gas corresponding to this particular complex value of $\sqrt{\kappa}$ does exist,

- 1) it is no longer a simple traveling wave as the ion disturbances no longer obey a law of the form

$$\text{Exp} \left\{ \frac{2\pi i}{\lambda} (z - Vt) \right\};$$

- 2) the particle vibrations are very large compared with what they would be for a real value of $\sqrt{\kappa}$ (for if not, first-order wave theory approximations would apply).

As we are neglecting ν and as the rate of working of the external magnetic force on a moving particle is zero, 2) would imply a relatively large storing of energy by the medium. Since (for $U < 0$) there is a negative solution for $\sqrt{\kappa}$, and since the solution corresponding to (38) which we have so far left out, *viz.*,

$$\kappa - 1 + \frac{\tilde{\omega}_0^2}{\omega(\omega_H - \omega)} = - \frac{\left(1 - \frac{U}{C} \sqrt{\kappa}\right) \omega_0^2 / \omega^2}{\frac{\omega_H + \omega}{\omega} - \frac{U}{C} \sqrt{\kappa}} \quad (43)$$

gives, for all negative U , just one real value of $\sqrt{\kappa}$, which is also negative, it seems clear to the writer that the case under consideration corresponds to reflection of any energy which might originally have been traveling in the Oz direction, via either or both of the backward traveling waves (which are given by the above negative permitted values of $\sqrt{\kappa}$). This interpretation is quite consistent with the usual way in which "reflection at a thick uniformly ionized layer at frequencies above the critical" is arrived at. For such a layer, ignoring the effect of the earth's field,

$$z > 0, \quad \kappa = 1 - \frac{\omega_0^2}{\omega(\omega + i\nu)} \quad (\nu \ll \omega)$$

$$z < 0, \quad \kappa = -1. \quad (44)$$

The case $\omega > \omega_0$ gives complex values of $\sqrt{\kappa}$ (which are similar in type to those under discussion above, being not strictly pure imaginary). There also results a negative real root $\sqrt{\kappa} = -1$ to provide a channel for the carrying of reflected energy. The universally accepted interpretation here is to say that the reflection is total with the forward part in $z > 0$ exponentially damped, the "growing wave" corresponding to the other complex root being rejected. It is, therefore, only consistent to adopt a similar interpretation above.

Doing this, it now follows from the above theory that, under suitable circumstances, wave propagation against a streaming gas can suffer sharp breaks as the frequency is varied (U/C fixed). See (42), for example, which, when $\epsilon\omega_H/\omega \ll 1$, is approximately

$$1 - \frac{\sqrt{2}}{8} (\epsilon\omega_H/\omega)^{1/2} < \left| \frac{U}{C} \right| \frac{\omega_0 \sqrt{\omega}}{(\omega_H - \omega)^{3/2}}$$

$$< 1 + \frac{\sqrt{2}}{8} (\epsilon\omega_H/\omega)^{1/2}. \quad (45)$$

Since records of whistling atmospherics exhibiting abrupt truncations are sometimes observed, the above may provide a possible theoretical explanation of the phenomenon.

The Propagation of Electromagnetic Waves in Ionized Gases

Part II—Propagation Along Stationary Columns

F. H. NORTHOVER†

THE general equation for the electric field E of an electromagnetic wave propagated through a stationary ionic distribution will be obtained from (32) of Part I, \bar{I} and \bar{I} being determined by (26) and (27) of that part.

In what follows, we make the assumption that the wave is almost entirely carried by the negatively charged particles. This is equivalent to assuming that the wave frequency is large compared with the proton gyro-frequency (see Appendix I). Consequently, we take g as effectively equal to \bar{I} , which by (32) of Part I implies

$$\text{div } \bar{I} = \frac{i\omega}{4\pi} \text{div } E \quad (1)$$

so that (26) of that part becomes

$$(i\omega - \nu)\bar{I} = -\frac{1}{4\pi}\bar{\omega}_0^2 E + \omega_H \bar{I} \times n - \frac{1}{4\pi}\{(\text{div } E)g - \bar{\beta}^2 \text{grad div } E\}. \quad (2)$$

This equation can be solved for \bar{I} and, then, taking $g \doteq \bar{I}$ again in (32), Part I, this yields an equation in E alone. At the frequencies we are considering, the term in g above is unimportant; we shall therefore neglect it [see (22), Part I, *et seq.*]. Eq. (2) therefore becomes

$$(i\omega - \nu)\bar{I} = -\frac{1}{4\pi}\bar{\omega}_0^2 \bar{A} + \omega_H \bar{I} \times n \quad (3)$$

where

$$\bar{A} = E - \frac{\bar{\beta}^2}{\bar{\omega}_0^2} \text{grad div } E. \quad (4)$$

The solution of this vector equation is

$$\bar{I} = \frac{i\omega}{4\pi} \bar{\alpha} \{ \bar{A} - \xi^2(n \cdot \bar{A})n - i\xi \bar{A} \times n \} \quad (5a)$$

where

$$\bar{\alpha} = \bar{\sigma}^2 \xi \gamma / (1 - \xi^2); \quad \bar{\sigma} = \bar{\omega}_0 / \omega; \quad \xi = \omega_H / (\omega + i\nu) \\ \gamma = \omega / \omega_H. \quad (5b)$$

The equation satisfied by E is therefore [from (32), Part I]

$$\text{curl curl } E + k^2 \{ \bar{\alpha} [\bar{A} - \xi^2(n \cdot \bar{A})n - i\xi \bar{A} \times n] - E \} = 0 \quad (6a)$$

where

$$k = \omega / C. \quad (6b)$$

Note: It is possible to obtain, in a similar fashion, an equation in E in the case of a *simple* moving gas with no fixed background gas; for, in this case, we can similarly write $g \doteq I$ (neglecting the effect of the proton gas). Div I is then expressible in terms of div E , etc. The result is in the form (6a) with A replaced by A , where,

$$A = \bar{A} + \frac{i}{\omega} U \times \text{curl } E - \frac{1}{\omega^2} \left(i\omega - \frac{i\omega}{C} \sqrt{\kappa} U - \nu \right) (\text{div } E) U.$$

I. STANDARD-TYPE PROPAGATION DOWN NONHOMOGENEOUS COLUMNS

The radial distribution of the ionic density will be taken in the general form

$$\bar{N} = \bar{N}_I + (\bar{N}_e - \bar{N}_I)f(r/a) \quad (7)$$

where r is the distance from the axis of any point of the distribution, $f(0) = 1, f'(0) = 0, f' < 0$, otherwise: $f(\infty) = 0$. \bar{N}_e is the value of \bar{N} at the axis and $\bar{n} \rightarrow \bar{n}_I$ as $r \rightarrow \infty$.

The length a will be a measure of the "extent" or "thickness" of the irregularity; as $a \rightarrow \infty$, $d\bar{N}/dr \rightarrow 0$ for every r . Hence, according to the definition of standard-type propagation (see Part I), we require a solution for E which approaches the simple plane traveling wave (the components being Cartesian).

$$E_\infty = (1, i, 0) \exp \left(\frac{\nu\omega}{C} \sqrt{\kappa z} - i\omega t \right) \quad (8)$$

as $a \rightarrow \infty$, where, since $\lim_{a \rightarrow \infty} \bar{n} = \bar{n}_e$, κ takes the value appropriate to plane-wave propagation in a homogeneous medium $\bar{N} = \bar{N}_e$; viz.,

$$\kappa = 1 - \bar{\alpha}_e - \bar{\alpha}_e \xi \quad (9)$$

the suffix c denoting the value of $\bar{\alpha}$ when $\bar{N} = \bar{N}_e$.

The proper type of solution to seek for "standard-type" propagation, is thus

$$E = F \exp \left\{ \frac{i\omega}{C} \sqrt{\kappa z} + i\phi - i\omega t \right\} \quad (10)$$

where the cylindrical components of the vector F are

† Dept. of Mathematics, Carleton University, Ottawa, Can.

functions of r only which tend to 1, i , and 0, respectively, when $a \rightarrow \infty$, and κ takes the value (9).

In classical magneto-ionic theory the pressure term (*i.e.*, that involving $\bar{\beta}$) is usually neglected. This does not cause any error when only plane waves parallel to the applied magnetic field in homogeneous media are considered, as was seen in Part I, Section III.

When propagation down columns is considered, however, the term in $\bar{\beta}$ enters. Owing to the complexity which is introduced by retention of this term we shall proceed tentatively: first, we shall discuss the propagation when the pressure term is neglected; second, we shall try to elucidate the circumstances under which this approximation can be made; and, last, we shall try to extend the analysis to cover the $\bar{\beta}$ term.

Neglecting, for the time being, therefore, the term in $\bar{\beta}$, *i.e.*, taking $\bar{\Lambda}$ as effectively equal to \bar{E} , we find, on working out the components of curl curl \mathbf{E} in terms of \mathbf{F} and substituting in (6a), the following equations for F_1 , F_2 , F_3 :

$$\left. \begin{aligned} ik\sqrt{\kappa}F_3' + i\tau^{-1}F_2' + \tau^{-2}(F_1 + iF_2) \\ + k^2\{(\kappa - 1 + \bar{\alpha})F_1 - i\bar{\alpha}\xi F_2\} &= 0 \\ -F_2'' - \tau^{-1}(k\sqrt{\kappa}F_3 + F_2' - iF_1') \\ + \tau^{-2}(F_2 - iF_1) + k^2\{(\kappa - 1 + \bar{\alpha})F_2 + i\bar{\alpha}\xi F_1\} &= 0 \\ -F_3'' + \tau^{-1}(ik\sqrt{\kappa}F_1 - k\sqrt{\kappa}F_2 - F_3') \\ + \tau^{-2}F_3 + ik\sqrt{\kappa}F_1' + k^2(\bar{\sigma}^2\xi\gamma - 1)F_3 &= 0 \end{aligned} \right\} \quad (11)$$

the dash denoting $d/d\tau$.

The distribution is expressed as a function of r/a , not r , and we have to satisfy the conditions $F \rightarrow (1, i, 0)$ as $a \rightarrow \infty$, for every fixed r . Writing, therefore,

$$\tau = r/a \quad (12a)$$

and also

$$\bar{\alpha} = \bar{\alpha}_0 - \Delta\bar{\alpha}, \quad \bar{\sigma}^2 = \bar{\sigma}_0^2 + \bar{\delta} \quad (12b)$$

the equations become

$$\left. \begin{aligned} ika\sqrt{\kappa}F_3' + i\tau^{-1}F_2' + \tau^{-2}(F_1 + iF_2) \\ + k^2a^2\{\bar{\alpha}_0\xi(-F_1 - iF_2) - \Delta\bar{\alpha}(F_1 - i\xi F_2)\} &= 0 \\ -F_2'' - \tau^{-1}(ka\sqrt{\kappa}F_3 + F_2' - iF_1') + \tau^{-2}(F_2 - iF_1) \\ + k^2a^2\{\bar{\alpha}_0\xi(-F_1 + iF_1) - \Delta\bar{\alpha}(F_2 + i\xi F_1)\} &= 0 \\ -F_3'' + \tau^{-1}(ika\sqrt{\kappa}F_1 - ka\sqrt{\kappa}F_2 - F_3') + \tau^{-2}F_3 \\ + ika\sqrt{\kappa}F_1' + k^2a^2(\bar{\sigma}_0^2\xi\gamma - 1 + \xi\gamma\bar{\delta})F_3 &= 0 \end{aligned} \right\} \quad (13)$$

where the dash now denotes $d/d\tau$.

A. The Field at Long Distances from the Axis

When τ is large enough, we can find an asymptotic approximation of the type

$$F \sim e^{iK\kappa a\tau} \sum_{n=0} a_n \tau^{-u-n} \quad (14)$$

it being assumed that $f(\tau)$ is expansible in powers of $1/\tau$. Writing $a_n = (A_n, B_n, C_n)$, we have, from the highest order terms,

$$\left. \begin{aligned} -K\sqrt{\kappa}C_0 - (\bar{\alpha}_0\xi + \bar{\Delta})A_0 - (\bar{\alpha}_0 - \bar{\Delta})i\xi B_0 &= 0 \\ \{K^2 - (\bar{\alpha}_0\xi + \bar{\Delta})\}B_0 + (\bar{\alpha}_0 - \bar{\Delta})i\xi A_0 &= 0 \\ (K^2 + \bar{\sigma}_I^2\xi\gamma - 1)C_0 - K\sqrt{\kappa}A_0 &= 0 \end{aligned} \right\} \quad (15)$$

Therefore, K satisfies the quartic equation

$$\begin{aligned} K^4(\kappa + \bar{\alpha}_0\xi + \bar{\Delta}) - K^2[(\bar{\Delta} + \bar{\alpha}_0\xi)(\kappa + 1 - \bar{\sigma}_I^2\xi\gamma) \\ + \bar{\Delta}(1 + \xi)\{2\bar{\alpha}_0\xi + \bar{\Delta}(1 - \xi)\}] \\ - (\bar{\sigma}_I^2\xi\gamma - 1)(1 + \xi)\{2\bar{\alpha}_0\xi + \bar{\Delta}(1 - \xi)\}\bar{\Delta} = 0. \end{aligned} \quad (16)$$

By considering in detail the equations for the higher coefficients (A_i, B_i, C_i) $i \geq 1$ it appears that $u = \frac{1}{2}$ (as might have been expected from energy considerations) and, clearly, only roots of this quartic in K for which $\text{Im}(K) \geq 0$ are physically admissible.

The asymptotic approximation is therefore of the form

$$F \sim e^{iK_1\kappa a\tau} \sum_{n=0} a_n^{(1)} \tau^{-(1/2)-n} + e^{iK_2\kappa a\tau} \sum_{n=0} a_n^{(2)} \tau^{-(1/2)-n} \quad (17)$$

where K_1 and K_2 are the two allowable roots of the quartic.

Substituting the values of the parameters in terms of the various frequencies and neglecting the collision frequency

$$\begin{aligned} \{1 - \bar{\Lambda} + \bar{G}(1 - \gamma^2)\}K^4 + \frac{1}{\bar{G}\gamma^3(1 - \gamma)} \\ \cdot \{2\gamma - 1 + \bar{\Lambda}(2\gamma^2 - 4\gamma + 1) + 2\bar{\Lambda}^2\gamma(1 - \gamma) \\ + 2\bar{G}\gamma^2(1 - \gamma)(1 + \bar{\Lambda}\gamma)\}K^2 \\ - \frac{1}{\bar{G}^2\gamma^4(1 - \gamma)}(1 - \bar{\Lambda} - \bar{G}\gamma^2)\{2 - \bar{\Lambda}(1 - \gamma)\}\bar{\Lambda} = 0, \end{aligned} \quad (18)$$

where

$$\left. \begin{aligned} \bar{\Lambda} &= (\bar{N}_e - \bar{N}_I)/\bar{N}_e \\ \gamma &= \omega/\omega_H \text{ (as already defined)} \\ \bar{G} &= \omega_H^2/\bar{\omega}_e^2. \end{aligned} \right\} \quad (19)$$

This equation can be written in the equivalent forms

$$\begin{aligned} \left\{1 + \frac{\bar{G}(1 - \gamma^2)}{1 - \bar{\Lambda}}\right\}K^4 + \frac{1}{\bar{G}\gamma^3(1 - \gamma)} \\ \cdot \left\{2\gamma - 1 - 2\bar{\Lambda}\gamma(1 - \gamma) + \frac{2\bar{G}\gamma^2(1 - \gamma)(1 + \bar{\Lambda}\gamma)}{1 - \bar{\Lambda}}\right\}K^2 \\ - \frac{1}{\bar{G}^2} \frac{2 - \bar{\Lambda}(1 - \gamma)}{\gamma^4(1 - \gamma)} \left(1 - \frac{\bar{G}\gamma^2}{1 - \bar{\Lambda}}\right)\bar{\Lambda} = 0. \end{aligned} \quad (20)$$

$$\begin{aligned} \left\{1 + \frac{\bar{G}(1 - \gamma^2)}{1 - \bar{\Lambda}}\right\}K^4 + \frac{1}{\bar{G}\gamma^3(1 - \gamma)} \\ \cdot \left\{1 + 2\left(\frac{\bar{G}\gamma^2}{1 - \bar{\Lambda}} - 1\right)(1 - \gamma)(1 + \bar{\Lambda}\gamma)\right\}K^2 \\ - \frac{1}{\bar{G}^2} \frac{2 - \bar{\Lambda}(1 - \gamma)}{\gamma^4(1 - \gamma)} \left(1 - \frac{\bar{G}\gamma^2}{1 - \bar{\Lambda}}\right)\bar{\Lambda} = 0. \end{aligned} \quad (21)$$

For strong guiding of the waves by the column, we require that:

- 1) the field should decrease radially at a much greater rate than does the radial scattering factor $r^{-1/2}$;
- 2) with condition 1) satisfied, the central regions of the column should carry a reasonable proportion of the total energy carried by the cylindrical distribution.

We are now in a position to examine the circumstances under which the fundamental condition 1) holds. To elucidate condition 2), we need to know the law of distribution of the column which (for example) affects the constant coefficients A_i , B_i , C_i in the above asymptotic expansion. It is, therefore, much more difficult to discuss than 1), prompting such a discussion to be postponed until later in the paper.

For condition 1) to be satisfied, we require *either* that both K_1^2 and K_2^2 be negative, *or* that both be complex. It must be noted that in connection with condition 1), $\bar{\Lambda}$ cannot, of course, be allowed to become too small, because, since one of the roots K_1 , K_2 tends to zero with $\bar{\Lambda}$, r would have to become very large before the radial exponential decline stipulated by condition 1) made itself felt. This point will appear again later. Essentially, $\bar{\Lambda} < 1$, and, in the cases in which we are interested (whistler propagation), $\gamma < 1$ also. *Condition 1) therefore requires that*

$$\{2 - \bar{\Lambda}(1 - \gamma)\} \left\{1 - \frac{\bar{G}\gamma^2}{1 - \bar{\Lambda}}\right\} \bar{\Lambda} < 0; \quad (22)$$

and either

$$a) \quad 1 + 2 \left(\frac{\bar{G}\gamma^2}{1 - \bar{\Lambda}} - 1 \right) (1 - \gamma)(1 + \bar{\Lambda}\gamma) > 0 \quad (23)$$

- b) or that the discriminant of the quadratic in K^2 should be negative, *i.e.* (after some reduction), that

$$\begin{aligned} &\{ (2\gamma - 1)^2 + 4\bar{\Lambda}\gamma(1 - \gamma) \} + 4\gamma^2(1 - \gamma) \\ &\cdot \{ (2\gamma - 1) + \bar{\Lambda}(1 - \gamma) \} \bar{G} + 4\gamma^4(1 - \gamma)^2 \bar{G}^2 < 0. \end{aligned} \quad (24)$$

If $\bar{\Lambda} > 0$, it appears that (24) can never be satisfied. This can also be seen from the alternative form (31) given below. Accordingly, the requirements reduce, for this case, to (22) and (23) only; *i.e.*, to the single condition.

$$\gamma > \{ (1 - \bar{\Lambda})/\bar{G} \}^{1/2} \quad (25)$$

which, since $\gamma < 1$ for whistler propagation, must necessarily entail the further condition

$$1 - \bar{\Lambda} < \bar{G}. \quad (26)$$

We turn now to a consideration of the columns for which $\bar{\Lambda} < 0$. In this case, the conditions (22) and (23) are satisfied if

$$\gamma_1 < \gamma < \{ (1 - \bar{\Lambda})/\bar{G} \}^{1/2} \quad (27a)$$

where γ_1 is the root between 0 and 1 of

$$1 + 2 \left(\frac{\bar{G}\gamma^2}{1 - \bar{\Lambda}} - 1 \right) (1 - \gamma)(1 + \bar{\Lambda}\gamma) = 0. \quad (27b)$$

For most of the range $0 < \gamma < 1$ to be covered, we must have first $\bar{G} < 1 - \bar{\Lambda}$ or \bar{G} near $1 - \bar{\Lambda}$; we then have the close approximation

$$\gamma_1 \approx \frac{1}{-2\bar{\Lambda}} \{ 1 - \bar{\Lambda} - \sqrt{1 + \bar{\Lambda}^2} \}, \quad (28)$$

for this makes $\bar{G}\gamma_1^2/(1 - \bar{\Lambda})$ certainly $< \frac{1}{4}(\bar{\Lambda} < 0)$.

Then, most of the range $0 < \gamma < 1$ will be covered provided that $-\bar{\Lambda} \gg 1$, for only then will γ_1 be small; *i.e.*, then

$$\gamma_1 \approx \frac{1}{-2\bar{\Lambda}} \quad (29)$$

for $-\bar{\Lambda} \gg 1$.

Considering still the case $\bar{\Lambda} < 0$, we now write down the conditions that (22) and (24) may be satisfied; for this we require that

$$\gamma < \{ (1 - \bar{\Lambda})/\bar{G} \}^{1/2} \quad (30)$$

and also [after a little reduction from (24)]

$$\bar{\Lambda} < \frac{-1}{1 - \gamma} \left(\frac{1}{4y\gamma} + y\gamma - 1 \right) \quad (31a)$$

where

$$y = 1 + \gamma(1 - \gamma)\bar{G}. \quad (31b)$$

It appears, as before, that a necessary condition for the fundamental condition (22) above to be satisfied, when $\bar{\Lambda} < 0$, over most of the frequency range given by $0 < \gamma < 1$ is that \bar{G} be either near $1 - \bar{\Lambda}$ or less than $1 - \bar{\Lambda}$. The necessary and sufficient condition that this be the case is therefore that

$$\max(\gamma_0, \gamma_1) \ll 1 \quad (32)$$

where γ_0 is the smaller of the two values of γ in $0 < \gamma < 1$ between which (31) is satisfied. By considering the unreduced form of the discriminant of (21) it is easy to see that $\gamma_1 > \gamma_0$; hence, by (28), (32) will only be satisfied if $-\bar{\Lambda} \gg 1$ and then there is a low-frequency cutoff given by

$$\gamma = \gamma_1 \approx 1/(-2\bar{\Lambda}). \quad (33)$$

II. SUMMARY OF THESE RESULTS

When $\bar{N}_e > \bar{N}_I(\bar{\Lambda} > 0)$, the necessary and sufficient condition that the field should decline exponentially in a radial direction outwards from the column at large distances from the axis is

- a) $\bar{N}_I/\bar{N}_e < f_H^2/\bar{f}_e^2$
- b) $(\bar{N}_I/\bar{N}_e)^{1/2}(\bar{f}_e/f_H) < \gamma < 1$

where f_H , \bar{f}_e are, respectively, the gyro-frequency and the axial plasma frequency. The cutoff frequency is here

$(\bar{N}_I/\bar{N}_e)^{1/2}\bar{f}_e$ so that for this guiding effect to hold over most of the frequency range $0 < f < f_H$, we must have

$$(\bar{N}_I/\bar{N}_e)^{1/2}(\bar{f}_e/f_H) \ll 1.$$

When $\bar{N}_e < \bar{N}_I$ ($\bar{\Lambda} < 0$) the necessary and sufficient condition that the field should decline exponentially in a radial direction outwards from the column at large distances from the axis is

$$\gamma_1 < \gamma < \min \{1; (\bar{N}_I/\bar{N}_e)^{1/2}(f_e/f_H)\}$$

where γ_1 is as defined by (28) and (29). Hence, for this guiding effect to hold over most of the frequency range $0 < f < f_H$, we must have

$$a) N_I/N_e > f_H^2/\bar{f}_e^2$$

or these two ratios approximately equal and,

$$b) N_I/N_e \gg 1.$$

A. The Field Near the Axis

As has been pointed out already, to have a rapidly decreasing field in a radial direction at sufficient distances from the axis is a necessary but not sufficient condition for the column to act as a good localized guide of electromagnetic energy. For this we require also that the central (axial) regions of the column carry a reasonable proportion of the total energy carried by the distribution.

We need, therefore, an estimate of the field near the central part of the column. As the preceding asymptotic-type approximation is inadequate for this, we shall develop a series-type approximation. In order to do this, we must suppose that the distribution function $f(\tau)$ is expansible in a series of powers of τ^2 , that is,

$$f(\tau) = 1 - f_1\tau^2 - f_2\tau^4 - \dots \quad (34)$$

As $a \rightarrow \infty$ we must have $F_1 \rightarrow 1$, $F_2 \rightarrow i$, $F_3 \rightarrow 0$. Also, it can be shown, from consideration of (13) for these quantities, that series expansions for F_1 and F_2 can contain only even powers of τ and a series for F_3 only odd powers. It is therefore assumed that

$$\begin{aligned} F_1 &= 1 + \lambda_1\tau^2 + \lambda_2\tau^4 + \dots + \lambda_n\tau^{2n} + \dots \\ F_2 &= i + \mu_1\tau^2 + \mu_2\tau^4 + \dots + \mu_n\tau^{2n} + \dots \\ F_3 &= \nu_0\tau + \nu_1\tau^3 + \dots + \nu_n\tau^{2n+1} + \dots \end{aligned} \quad (35)$$

For any r , $\tau \rightarrow 0$ as $a \rightarrow \infty$. The coefficients in the series tend to infinity with a but (as will be seen later and in Appendix II) they do not increase sufficiently rapidly with a to prevent every variable term in these series tending to zero as $a \rightarrow \infty$. Thus, the series satisfy the fundamental requirement above on F_1 , F_2 , F_3 .

Denoting by the abbreviation " $\bar{\Delta}$ " the value of $\Delta\bar{\alpha}$ at $r = \infty$, then

$$\bar{\Delta} = \bar{\alpha}_e - \bar{\alpha}_I \quad (36)$$

and

$$\begin{aligned} \alpha &= \bar{\alpha}_I + (\bar{\alpha}_e - \bar{\alpha}_I)f(\tau) \\ &= \bar{\alpha}_e - \bar{\Delta}\{1 - f(\tau)\} \end{aligned} \quad (37)$$

so that in (13) we use

$$\Delta\bar{\alpha} = \bar{\Delta}\{1 - f(\tau)\}$$

where $f(\tau)$ is expanded as in (34).

The components H_1 , H_2 , H_3 of the magnetic field in the wave are given in terms of F_1 , F_2 , F_3 , by

$$\left. \begin{aligned} ikH_1 &= (ir^{-1}F_3 - ik\sqrt{\kappa}F_2)e^{i\phi + ik\sqrt{\kappa}z} \\ ikH_2 &= (ik\sqrt{\kappa}F_1 - F_3')e^{i\phi + ik\sqrt{\kappa}z} \\ ikH_3 &= (1/r)\{(rF_2)' - iF_1\}e^{i\phi + ik\sqrt{\kappa}z} \end{aligned} \right\}. \quad (38)$$

For the standard plane wave, obtained when $a \rightarrow \infty$, i.e., when $\tau \rightarrow 0$,

$$\mathbf{E} = (1, i, 0)e^{i\phi + ik\sqrt{\kappa}z}. \quad (39)$$

The value of the magnetic field in this standard plane wave is easily found to be

$$\sqrt{\kappa}(-i, 1, 0)e^{i\phi + ik\sqrt{\kappa}z}. \quad (40)$$

Hence, letting $a \rightarrow \infty$ ($\tau \rightarrow 0$) and using the facts that $F_2 \rightarrow i$, $F_1 \rightarrow 1$, we have from the above, that $F_3' = 0$. Thus $\tau \rightarrow 0$.

Substituting the assumed series expressions in the three simultaneous linear second-order differential equations in (13), the following difference equations for the coefficients are obtained:

$$\begin{aligned} (2n+1)ika\sqrt{\kappa}\nu_n + (2n+3)i\mu_{n+1} + \lambda_{n+1} \\ + k^2a^2\{\bar{\alpha}_e\xi(-\lambda_n - i\mu_n) - \bar{\Delta}(\lambda_{n-1} - i\xi\mu_{n-1})f_1 \\ - \bar{\Delta}(\lambda_{n-2} - i\xi\mu_{n-2})f_2 - \dots\} = 0 \end{aligned} \quad (41a)$$

$$\begin{aligned} -ka\sqrt{\kappa}\nu_n - (2n+1)(2n+3)\mu_{n+1} + i(2n+1)\lambda_{n+1} \\ + k^2a^2\{\bar{\alpha}_e\xi(-\mu_n + i\lambda_n) - \bar{\Delta}(\mu_{n-1} + i\xi\lambda_{n-1})f_1 \\ - \bar{\Delta}(\mu_{n-2} + i\xi\lambda_{n-2})f_2 - \dots\} = 0 \end{aligned} \quad (41b)$$

$$\begin{aligned} -4(n+1)(n+2)\nu_{n+1} + k^2a^2(\bar{\sigma}_e^2\xi\gamma - 1)\nu_n \\ + k^2a^2\xi\gamma\bar{\delta}_e(f_1\nu_{n-1} + f_2\nu_{n-2} + \dots) \\ + ka\sqrt{\kappa}\{(2n+3)i\lambda_{n+1} - \mu_{n+1}\} = 0 \end{aligned} \quad (41c)$$

where

$$\bar{\delta}_e = \bar{\sigma}_e^2 - \bar{\sigma}_I^2. \quad (41d)$$

These equations give, after somewhat heavy algebra,

$$\left. \begin{aligned} \lambda_1 &= -\frac{3}{4} \frac{f_1(1+\xi)\bar{\Delta}}{1-\bar{\alpha}_e} \\ \mu_1 &= -\frac{i}{4} \frac{f_1(1+\xi)}{1-\bar{\alpha}_e} \bar{\Delta} \\ \nu_1 &= -\frac{ika\sqrt{\kappa}}{4} \frac{f_1(1+\xi)}{1-\bar{\alpha}_e} \end{aligned} \right\}. \quad (42a)$$

$$\begin{aligned}
24\lambda_2 = & -\frac{5}{4} \left(\frac{k^2 a^2 \bar{\alpha}_c \xi (\bar{\sigma}_c^2 \xi \gamma - 1 + \kappa) - 2(7 + \xi) f_1 \bar{\Delta}_c}{(1 - \bar{\alpha}_c)^2} \right) f_1 \bar{\Delta} (1 + \xi) - \frac{20 f_2 (1 + \xi)}{1 - \bar{\alpha}_c} \bar{\Delta} - \frac{5}{2} k^2 a^2 \frac{(1 + \xi) f_1 \bar{\alpha}_c \xi \bar{\Delta}}{1 - \bar{\alpha}_c} \\
& + \frac{k^2 a^2}{4} \frac{f_1 \bar{\Delta} (1 + \xi)}{1 - \bar{\alpha}_c} \{ 5(\bar{\sigma}_c^2 \xi \gamma - 1) + \kappa \} - \frac{1}{2} k^2 a^2 \bar{\Delta} (1 + \xi) f_1 \\
24\mu_2 = & -\frac{i}{4} \left(\frac{k^2 a^2 \bar{\alpha}_c \xi (\bar{\sigma}_c^2 \xi \gamma - 1 - \kappa) - 2(7 + \xi) f_1 \bar{\Delta}}{(1 - \bar{\alpha}_c)^2} \right) f_1 \Delta (1 + \xi) + \frac{i k^2 a^2}{4} \frac{f_1 \bar{\Delta} (1 + \xi)}{1 - \bar{\alpha}_c} (\bar{\sigma}_c^2 \xi \gamma - 1 + 5\kappa) \\
& - \frac{5}{2} i k^2 a^2 \bar{\Delta} (1 + \xi) f_1 - \frac{4 i f_1 (1 + \xi) \bar{\Delta}}{1 - \bar{\alpha}_c} - \frac{i k^2 a^2 \bar{\Delta} (1 + \xi) f_1 \bar{\alpha}_c \xi}{2(1 - \bar{\alpha}_c)} \\
24\nu_2 = & -\frac{i k a \sqrt{\kappa}}{4} \left(\frac{k^2 a^2 \bar{\alpha}_c \xi (\bar{\sigma}_c^2 \xi \gamma - 1 - \kappa) - 2(7 + \xi) f_1 \bar{\Delta}}{(1 - \bar{\alpha}_c)^2} \right) f_1 \bar{\Delta} (1 + \xi) - \frac{4 i k a \sqrt{\kappa} f_2 (1 + \xi) \bar{\Delta}}{1 - \bar{\alpha}_c} \\
& - \frac{i k^3 a^3 \sqrt{\kappa}}{2} \frac{\bar{\Delta} (1 + \xi) f_1 \bar{\alpha}_c \xi}{1 - \bar{\alpha}_c}
\end{aligned} \quad (42b)$$

We shall later need to know the order of magnitude of λ_n , μ_n , and ν_n , when n is large and ka is large. We note that λ_2 , μ_2 are polynomials in a of degree 2 while ν_2 is a polynomial of degree 3 with factor a . The general result is that λ_n , μ_n are polynomials of degree $2n-2$ in a , while ν_n is a polynomial of degree $2n-1$ in a (containing at least a factor a). This can be proven by mathematical induction using the difference equations of (41) (see Appendix II). F_1 , F_2 and F_3 therefore tend to the required limits as $a \rightarrow \infty$.

B. The Energy Flow in the Axial Regions

Writing for brevity,

$$\psi = \phi + k\sqrt{\kappa}z - \omega t \quad (43)$$

and remembering that the actual physical field components are to be obtained by taking the real parts of the mathematical expressions in the text, there results

$$\left. \begin{aligned}
H_1 &= \text{Re} \left(\frac{F_3}{kr} - F_2 \sqrt{\kappa} \right) e^{i\psi} \\
H_2 &= \text{Re} \left(F_1 \sqrt{\kappa} + \frac{i F_3'}{k} \right) e^{i\psi} \\
H_3 &= \text{Re} \left\{ -\frac{i}{kr} (F_2 - i F_1 + r F_2') \right\} e^{i\psi}
\end{aligned} \right\}, \quad (44)$$

the dash denoting differentiation with respect to r . Likewise,

$$\left. \begin{aligned}
E_1 &= \text{Re} (F_1 e^{i\psi}) \\
E_2 &= \text{Re} (F_2 e^{i\psi}) \\
E_3 &= \text{Re} (F_3 e^{i\psi})
\end{aligned} \right\}. \quad (45)$$

Using the series which have been developed for F_1 , F_2 , and F_3 ,

$$\left. \begin{aligned}
H_1 &= - \left\{ \frac{1}{ka\tau} \left(\frac{\nu_1}{i} \tau^3 + \dots \right) - \sqrt{\kappa} \left(1 + \frac{\mu_1}{i} \tau^2 + \dots \right) \right\} \sin \psi \\
H_2 &= \left\{ \sqrt{\kappa} (1 + \lambda_1 \tau^2 + \dots) - \frac{1}{ka} \left(\frac{3\nu_1}{i} \tau^2 + \dots \right) \right\} \cos \psi \\
H_3 &= 0(\tau^3) \cos \psi
\end{aligned} \right\} \quad (46)$$

$$\left. \begin{aligned}
E_1 &= (1 + \lambda_1 \tau^2 + \dots) \cos \psi \\
E_2 &= - \left(1 + \frac{\mu_1}{i} \tau^2 + \dots \right) \sin \psi \\
E_3 &= - \left(\frac{\nu_1}{i} \tau^3 + \dots \right) \sin \psi
\end{aligned} \right\}. \quad (47)$$

Whence

$$(E \wedge H)_1 = 0(\tau^3) \sin \psi \cos \psi. \quad (48)$$

$$(E \wedge H)_3 = \sqrt{\kappa} \{ 1 + \lambda_1 (\cos^2 \psi + \frac{1}{3} \sin^2 \psi) \tau^2 + \dots \}. \quad (49)$$

The flux of energy is therefore directed wholly parallel to the axis of the distribution, since the mean value of $(E \wedge H)_1$ over a time cycle is zero.

The total flux of energy down the column is thus

$$\begin{aligned}
& \frac{Ca^2}{4\pi} \int_0^\infty \int_0^{2\pi} \tau d\tau d\phi (E \wedge H)_3 \\
&= \frac{Ca^2}{2} \int_0^\infty \tau d\tau \sqrt{\kappa} \{ 1 + b_1 \tau^2 + b_2 \tau^4 + \dots \} \quad (50)
\end{aligned}$$

by the above, where

$$\left. \begin{aligned} b_1 &= \lambda_1 + \frac{\mu_1}{i} - \frac{2\nu_1}{ika\sqrt{\kappa}} \\ b_2 &= \frac{1}{2}(\lambda_1^2 - \mu_1^2) + \left(\lambda_2 + \frac{\mu_2}{i}\right) - \frac{3\nu_2}{ika\sqrt{\kappa}} \\ &\quad - \frac{\nu_1}{3ika\sqrt{\kappa}} \left(3\lambda_1 + \frac{\mu_1}{i}\right) \end{aligned} \right\} \quad (51)$$

We must now make an estimate of the order of magnitude of this integral when the necessary condition 1) of Section II-A, for good guiding over most of the range, $0 < \gamma < 1$, is satisfied.

C. $\bar{\Lambda} \ll -1$; Permitted F-Range Effectively

$$\frac{1}{2}f_H/(-\bar{\Lambda}) < f < f_H$$

As seen, the satisfaction of the necessary condition 1) of Section II-A implies either that \bar{G} be near $1 - \bar{\Lambda}$ or that $\bar{G} < 1 - \bar{\Lambda}$. It is not likely that \bar{G} would ever be much larger than unity in applications to whistlers and, since

provided that γ is not too near the upper or lower limits referred to above. Hence, with this proviso,

$$ika\mathcal{G}(K) = -2\pi X(-\bar{\Lambda})^{1/2} \quad (55)$$

where

$$X = a/\lambda_c \quad (56)$$

and λ_c , the wavelength of the plasma frequency at the axis, is given by

$$\bar{\omega}_c/C = 2\pi/\lambda_c. \quad (57)$$

Thus, owing to the appearance of the exponential factor in the field at sufficient radial distances, the contribution of the range $\frac{1}{2}X^{-1}(-\bar{\Lambda})^{-1/2}$ to infinity, to the integral (50), can be neglected in an approximate appraisal of its value.

After some heavy algebra, we find the following reduced forms for the second order coefficients λ_2 , μ_2 , ν_2 , which occur in the series for F_1 , F_2 , F_3

$$\left. \begin{aligned} \lambda_2 &= \left\{ -\frac{\pi^2 X^2}{6} \frac{5 + 5\gamma + 6\gamma^2 - 2\bar{G}\gamma^2(1 + \gamma)(5 + 3\gamma)}{\gamma(1 + \gamma)} + \frac{5}{36} \left(\frac{1 + 7\gamma}{\gamma} \frac{f_1 \bar{\Lambda}}{1 + \bar{G}(1 - \gamma^2)} + \frac{8f_2}{f_1} \right) \right\} \lambda_1 \\ \mu_2 &= \left\{ -\frac{\pi^2 X^2}{6} \frac{1 + \gamma + 6\gamma^2 - 2\bar{G}\gamma^2(1 + \gamma)(1 + 3\gamma)}{\gamma(1 + \gamma)} + \frac{1}{12} \left(\frac{1 + 7\gamma}{\gamma} \frac{f_1 \bar{\Lambda}}{1 + \bar{G}(1 - \gamma^2)} + \frac{8f_2}{f_1} \right) \right\} \mu_1 \\ \nu_2 &= \left\{ -\frac{\pi^2 X^2}{6} \frac{1 - 2\bar{G}\gamma^2(1 + \gamma)}{\gamma(1 + \gamma)} + \frac{1}{12} \left(\frac{1 + 7\gamma}{\gamma} \frac{f_1 \bar{\Lambda}}{1 + \bar{G}(1 - \gamma^2)} + \frac{8f_2}{f_1} \right) \right\} \nu_1 \end{aligned} \right\} \quad (58)$$

After further algebra, we find that the series in the integrand of (50) is

$$\begin{aligned} 1 + \frac{2}{3} \lambda_1 \tau^2 - \lambda_1 \tau^4 &\left\{ -\frac{\pi^2 X^2}{18} \frac{13 + 16\gamma + 24\gamma^2 + 8\bar{G}\gamma(1 + \gamma)(2 + \gamma - 6\gamma^2) - 24\bar{G}^2\gamma^3(1 + \gamma)^2(1 - \gamma)}{\{1 + \bar{G}(1 - \gamma^2)\}\gamma(1 + \gamma)} \right. \\ &\quad \left. + \frac{1}{12} \left(\frac{1 + 7\gamma}{\gamma} \frac{f_1 \bar{\Lambda}}{1 + \bar{G}(1 - \gamma^2)} + \frac{8f_2}{f_1} \right) \right\}. \end{aligned} \quad (59)$$

we have to have $-\bar{\Lambda} \gg 1$ as a necessary condition for propagation over most of the frequency range $0 < f < f_H$, (21) becomes, for this case, effectively,

$$\begin{aligned} K^4 + \frac{1}{\bar{G}\gamma^3(1 - \gamma)} \{2\gamma - 1 - 2\bar{\Lambda}\gamma(1 - \gamma)\} K^2 \\ - \frac{1}{\bar{G}^2\gamma^4} \frac{2 - \bar{\Lambda}(1 - \gamma)}{1 - \gamma} \bar{\Lambda} = 0, \end{aligned} \quad (52)$$

where (since $-\bar{\Lambda} \gg 1$)

$$K^2 = \frac{\bar{\Lambda}}{\bar{G}\gamma^2} \left\{ 1 \pm \frac{1}{\sqrt{\bar{\Lambda}\gamma(1 - \gamma)}} + 0\{[\bar{\Lambda}\gamma(1 - \gamma)]^{-1}\} \right\} \quad (53)$$

that is,

$$K = \frac{i}{\gamma^2} \left(\frac{-\bar{\Lambda}}{\bar{G}} \right)^{1/2}, \quad (54)$$

Writing τ_m equal to the new upper limit $\frac{1}{2}X^{-1}(-\bar{\Lambda})^{-1/2}$ of the integral in (50), when approximated, as explained above, and considering the order of magnitude of the first few terms of the series, we have, for example,

$$\begin{aligned} \lambda_1 \tau_m^2 &= \frac{3}{4} \frac{f_1(1 + \gamma)}{\gamma} \frac{\bar{\Lambda}}{1 + \bar{G}(1 - \gamma^2)} \frac{1}{4X^2 |\bar{\Lambda}|} \\ &= 0 \left(\frac{1}{X^2 \gamma} \right) < 0(|\bar{\Lambda}|/X^2). \end{aligned} \quad (60)$$

If $X^2/|\bar{\Lambda}| \gg 1$ or order unity, then this is small or order unity and the order of the fourth-degree term is then

$$X^2 |\bar{\Lambda}| \gamma^{-2} \tau_m^4 = \frac{1}{X^2 \gamma^2 |\bar{\Lambda}|} < 0(|\bar{\Lambda}|/X^2). \quad (61)$$

Considering, therefore, first the case $X^2 \gg |\bar{\Lambda}|$ or $X^2 = 0(|\bar{\Lambda}|)$ then, if $X^2 \gg |\bar{\Lambda}|$, the flux of energy down the column approximates to

$$\frac{1}{4} C \lambda_c^2 \chi \quad (62)$$

where

$$\begin{aligned} \chi &= X^2 \sqrt{\kappa \tau_m^2} \\ &= \frac{1}{4} \frac{\bar{\omega}_c}{\omega_H} \frac{1}{|\bar{\Lambda}| \{\gamma(1-\gamma)\}^{1/2}} \end{aligned} \quad (63)$$

If $X^2 = 0(|\bar{\Lambda}|)$, the energy flux is of this order of magnitude. The flux of energy of the unit field $E(1, i, 0)$ down a column of the "standard" radius λ_c is here taken as our standard of comparison for measurement of energy flux. χ is then the ratio of the actual flux to this standard flux.

We have now, however, to consider the case $X^2 \ll |\bar{\Lambda}|$; the difficulty here is that the series in question cannot be replaced by its first few terms. However, it will be noticed that the coefficient of τ^4 in the F series (and so in the above series) is a polynomial in $\bar{\Lambda}/\gamma^2$ of degree 2. The coefficients of this polynomial are of course independent of $\bar{\Lambda}$ but are bounded, continuous functions of \bar{G} and γ throughout $0 \leq \gamma \leq 1$, $\bar{G} \geq 0$. By mathematical induction (see Appendix II), the general result for the coefficient of τ^{2n} can be reached; hence, the above series, for large $|\bar{\Lambda}|$ and $X^2 \ll |\bar{\Lambda}|$ is of the form

$$1 + a_1 \frac{\bar{\Lambda}}{\gamma} \tau^2 + a_2 \frac{\bar{\Lambda}^2}{\gamma^2} \tau^4 + \dots + 0 \left(\frac{X^2}{\bar{\Lambda}} \right) + 0 \left(\frac{1}{\bar{\Lambda}} \right) \quad (64)$$

where the a 's are continuous, bounded functions of γ and \bar{G} for $0 \leq \gamma \leq 1$, $\bar{G} \geq 0$.

Although this series may diverge for sufficiently large τ [it does, for example, if the distribution law is $(1+\tau^2)^{-1}$], the function which it represents is analytic along the whole real τ axis (since the field is known to be finite and continuous for all r), and tends to zero as $\tau \rightarrow +\infty$. Denoting this function by $\Phi(\bar{G}, \gamma, \theta)$ where $\tau^2 = \theta\gamma/\bar{\Lambda}$, then (since the function is derived directly from the field functions F_1, F_2, F_3 which are continuous with continuous first derivatives), it follows that

$$d\Phi(\bar{G}, \gamma, \theta)/d\theta$$

is continuous in θ and also in \bar{G}, γ for any fixed θ . Hence, θ_m (where θ_m gives the maximum value of Φ for θ -variations) is continuous in \bar{G}, γ . Hence, $\Phi(\bar{G}, \gamma, \theta_m)$ is continuous in \bar{G}, γ and is therefore bounded.

Hence, in this case also, the order of the integral (50) is as found above. If, however, X is so small that $\tau_m > 1$, the field spreads appreciably away from the axial regions. This happens if $X < |\bar{\Lambda}|^{-1/2}$; we must not have this for good guiding.

Hence, in the case under consideration ($-\bar{\Lambda} \gg 1$) we obtain good guiding throughout most of the frequency range

$$\frac{1}{4} f_H / (-\bar{\Lambda}) < f < f_H,$$

when $(\bar{f}_c/f_H)(\bar{N}_c/\bar{N}_I)$ is as large as possible (e.g., appreciably greater than unity) subject to the requirements $(\bar{N}_I/\bar{N}_c) \gg 1$; $X > \frac{1}{2}(\bar{N}_c/\bar{N}_I)^{1/2}$, e.g. Obviously, these requirements cannot be met unless f_H/\bar{f}_c is small. Alternatively, we may state the requirements as follows. Writing the (small) lower cutoff frequency as $\gamma_L f_H$ (so that $\gamma_L = \frac{1}{2}(\bar{N}_c/\bar{N}_I)$), we require that $2(\bar{f}_c/f_H)\gamma_L$ should be as large as possible (appreciably greater than unity, e.g.) and also that the column satisfy the "thickness condition" $X > \sqrt{\gamma_L}$.

D. The Case $0 < 1 - \bar{\Lambda} \ll 1$; Permitted F-Range $\{(1 - \bar{\Lambda})/\bar{G}\}^{1/2} f_H < f < f_H$

In this case, it is essential that

$$1 - \bar{\Lambda} \ll \bar{G} \quad (65)$$

i.e., that

$$(N_c/N_I) \gg (\bar{f}_c/f_H)^2 \quad (66)$$

for the frequency range $0 < f < f_H$ to be well covered. The equation for K is now, effectively,

$$\begin{aligned} &\bar{G}(1 - \gamma^2)K^4 \\ &+ \frac{2(1 + \bar{\Lambda}\gamma)}{\gamma} K^2 + \frac{1}{\bar{G}\gamma^2} \frac{2 - \bar{\Lambda}(1 - \gamma)}{1 - \gamma} \bar{\Lambda} = 0, \end{aligned} \quad (67)$$

where

$$K^2 = -\frac{2 - \bar{\Lambda}(1 - \gamma)}{\gamma\bar{G}(1 - \gamma^2)} \text{ or } -\frac{\bar{\Lambda}}{\gamma\bar{G}(1 - \gamma)}. \quad (68)$$

Since $\bar{\Lambda}$ is nearly unity in the present case, we have, effectively,

$$K_1 = K_2 = \frac{i}{\sqrt{\bar{G}\gamma(1 - \gamma)}}. \quad (69)$$

Thus

$$ka\mathcal{G}(K) \doteq 2\pi X \sqrt{\frac{\gamma}{1 - \gamma}}. \quad (70)$$

Hence, here

$$\tau_m = \frac{1}{2} X^{-1} \sqrt{\frac{1 - \gamma}{\gamma}}, \quad (71)$$

and, for the field to be fairly well localized near the axis of the column, we require $\tau_m \leq 1$ for all the values of γ in the permitted range [this in view of the radial attenuation factor $ka\tau\mathcal{G}(K)$]; i.e., we require that

$$X^{-1} \{\bar{G}/(1 - \bar{\Lambda})\}^{1/4} < 2$$

i.e.,

$$X > \frac{1}{2} \{\bar{G}/(1 - \bar{\Lambda})\}^{1/4} \quad (72)$$

i.e.,

$$X > \frac{1}{2\sqrt{\gamma_L}} \quad (73)$$

where γ_L is the lower limit $\{(1-\bar{\Delta})/\bar{G}\}^{1/2}$ of the permitted range for γ . Hence, X has, in the present case, to be fairly large. Here, $X^2 \gg |\bar{\Delta}|$ and, considering the size of the first few terms of the series in (50),

$$\lambda_1 \tau_m^2 = 0(X^2 \gamma^{-2}) \quad (74)$$

and the term in τ^4 with $\tau = \tau_m$ is

$$0\left(\frac{X^2}{\gamma^2} \frac{X^{-4}}{\gamma^2}\right) = 0(X^{-2} \gamma^{-4}). \quad (75)$$

Hence, if γ is not too small, the series in (50) approximates to unity and then, using the value of τ_m from (71),

$$\begin{aligned} \chi &= X^2 X^{-2} \frac{1 - \gamma \bar{f}_c}{\gamma} \frac{1}{f_H \gamma (1 - \gamma)^{1/2}} \\ &= \gamma^{-3/2} (1 - \gamma)^{1/2} (f_c/f_H). \end{aligned} \quad (76)$$

when $\gamma \gg X^{-1/2}$. When γ becomes near this value χ is in general of this order of magnitude.

If γ becomes small (although of course still $> \gamma_L$), then the series in (50) can no longer be approximated by unity nor even by its first few terms. We then wish to know its order of magnitude.

Now since [by condition (73)], $X^2 \gg 1 \approx \bar{\Delta}$, and since, by Appendix II, the dominant part of the coefficient of τ^{2n} in the series in question is X^{2n-2}/γ^n times a function of γ , \bar{G} bounded throughout $0 \leq \gamma \leq 1$, $\bar{G} \geq 0$, then the series is of the form

$$1 + \frac{2}{3} \lambda_1 \tau^2 + \frac{X^2}{\gamma^2} \tau^4 P \quad (77a)$$

where P is a series of the form

$$\sum_{n=0}^{\infty} p_n \frac{X^{2n}}{\gamma^n} \tau^{2n} \quad (77b)$$

wherein the p 's are bounded functions of γ , \bar{G} and are independent of X .

As before, this represents a function whose continuation along the real axis is analytic, so that $P(X\tau/\sqrt{\gamma})$ is bounded for all real τ , tending to zero as $\tau \rightarrow \infty$ (because the complete series, being derived directly from F_1, F_2, F_3 , does).

Hence, the series in question is, when $\tau = \tau_m$ and γ is so small that it does not approximate to unity, of order

$$X^{-2} \gamma^{-4} \quad (78)$$

when $\gamma \leq X^{-1/2}$.

The order of χ is now

$$\begin{aligned} X^2 X^{-2} \frac{1 - \gamma \bar{f}_c}{\gamma} \frac{1}{f_H \sqrt{\gamma(1 - \gamma)}} X^{-2} \gamma^{-4} (1 - \gamma)^2 \\ = 0 \left\{ \frac{\bar{f}_c}{f_H} X^{-2} \gamma^{-1/2} (1 - \gamma)^{5/2} \right\} \quad (\gamma \leq X^{-1/2}). \end{aligned} \quad (79)$$

This estimate of the order of magnitude of χ is valid when $\gamma \leq X^{-1/2}$ [$\gamma > \frac{1}{4} X^{-2}$; see condition (71) with $\tau_m \leq 1$].

III. CONCLUSION

Energy can be carried either by columns of the central ionic surplus type (which we shall call S -columns) or by the central ionic deficiency type (D -columns) over a large part of the frequency range $0 < f < f_H$, provided that both types are well developed ($N_e \ll N_I$ or $N_e \ll N_L$) and are not too thin.

The main difference in the two cases is in the order of magnitude of the energy parameter χ ; apart from simple factors in γ , there is an extra small factor $|\bar{\Delta}|^{-1}$ in the expression for D -columns [see (76) and (63)]. It is therefore clear that, when the thickness parameter X is large enough for both types of column to be operative, far more energy is carried down an S -column than down a D -column. There are, however, some qualifying conditions over and above the necessity condition that the columns should be well developed. These are now explained.

If the S -columns are operative, they are much better energy carriers than the D -columns, but they have to be rather thick ($X > \frac{1}{2} \gamma_L^{-1/2}$). On the other hand, quite thin D -columns are permissible ($X > \sqrt{\gamma_L}$). However, since for them $-\bar{\Delta}$ cannot much exceed f_c/f_H (otherwise the carried energy becomes small) it follows that $\gamma_L \{ \approx 1/(-4\bar{\Delta}) \}$ cannot be much less than $\frac{1}{4} (f_H/\bar{f}_c)$. This may not allow a low enough cutoff frequency ($\gamma_L f_H$) to enable propagation to be explained in terms of this type of column. In view of this, therefore, and in view of the fact that the energy flux along S -columns is so much greater than along D -columns, it is likely that the whistler propagation can be accounted for by supposing the existence of columns of the former type.

A. The Effects of Gas Pressure

In the above analysis, we have neglected this effect; this [see (4)] is equivalent to making the assumption

$$\frac{\bar{\beta}^2}{\bar{\omega}_0^2} |\text{grad div } \mathbf{E}| \ll |\mathbf{E}| \quad (80)$$

$\bar{\beta}$ being the velocity of sound in a nonelectrified gas of equivalent pressure and density; i.e., $\bar{\beta} = (dp/d\rho)_0$. In this brief section is investigated the problem of how far this assumption is valid. For the components of $\text{grad div } \mathbf{E}$, it is found that

$$\begin{aligned} \text{grad div } \mathbf{E} = \left\{ \frac{d}{dt}; i\tau^{-1}; ika\sqrt{\kappa} \right\} a^{-2} \\ \cdot \{ F_1 + \tau^{-1}(F_1 + iF_2) + ika\sqrt{\kappa}F_3 \} e^{i\psi}. \end{aligned}$$

We shall investigate the order of magnitude of this: 1) near the axis, 2) far from the axis. It has, of course, to be remembered that, when taking the actual vectors, we have to take the real part of the analytical expressions.

$$1) \tau \ll 1:$$

$$\text{grad div } \mathbf{E} \approx a^{-2} (3\lambda_1 + i\mu_1) \{ 1, i, ika\sqrt{\kappa}\tau \} e^{i\psi}.$$

Using the effective expression $\bar{\omega}_0^2/\{\omega(\omega_H - \omega)\}$ for κ ,

$$ka\sqrt{\kappa} = \pi/\tau_m \quad (\bar{\Lambda} > 0) \\ = \pi \left\{ |\bar{\Lambda}|^{-1/2} \left(\frac{\gamma}{1-\gamma} \right)^{1/2} \right\} / \tau_m \quad (\bar{\Lambda} < 0).$$

Hence, (80) will be satisfied provided that

$$\bar{\beta}^2 a^{-2} \bar{\omega}_0^{-2} |3\lambda_1 + i\mu_1| \ll 1$$

i.e., provided that

$$\frac{1}{2\pi^2} \frac{|\bar{\Lambda}|(1+\gamma)}{X^2\gamma} \left(\frac{\bar{\beta}}{C} \right)^2 \ll 1.$$

Taking account of permitted ranges for γ and X in the two types of columns, satisfaction of (80) over these ranges requires

$$\bar{\beta}/C \ll |\bar{\Lambda}|^{-3/2} \quad (\bar{\Lambda} < 0) \\ \ll 1 \quad (\bar{\Lambda} > 0).$$

2) τ Large:

$$\text{grad div } \mathbf{E} \sim \left(\frac{d}{dt}; 0; ika\sqrt{\kappa} \right) a^{-2} (F_1' + ika\sqrt{\kappa}F_3)e^{i\psi} \\ \sim -k^2(K; 0; \sqrt{\kappa})(KF_1 + i\sqrt{\kappa}F_3)e^{i\psi}.$$

3) Case $\bar{\Lambda} > 0$: Here [see (69)] $K = \sqrt{\chi}$; hence, on taking real parts,

$$|\text{grad div } \mathbf{E}| \sim k^2\kappa \{F_1^2 + (F_3/i)^2\}^{1/2} \\ |\mathbf{E}| = [F_1^2 \cos^2 \psi + \{i(F_2 + F_3)\}^2 \sin^2 \psi]^{1/2}.$$

Hence, requiring

$$k^2\kappa = \frac{\bar{\beta}^2}{\bar{\omega}_0^2} \ll 1$$

i.e.,

$$\left(\frac{\bar{\beta}}{C} \right)^2 \frac{\gamma}{1-\gamma} \ll 1.$$

4) Case $\bar{\Lambda} < 0$: Here {see (54)}

$$K/\sqrt{\kappa} = i|\bar{\Lambda}|^{1/2}\{(1-\gamma)/\gamma\}^{1/2}.$$

Hence, except if γ is very near unity (i.e., provided that $1-\gamma \gg \bar{\Lambda}^{-1}$), the above dictates that the condition (80) is here

$$\frac{\bar{\beta}^2}{C^2} \frac{\omega}{\bar{\omega}_0} |\bar{\Lambda}|^{1/2} \ll 1.$$

In view of the probable smallness of $\bar{\beta}/C$, it seems likely that condition (80) will be satisfied; i.e., pressural effects can probably be neglected for column propagation.

APPENDIX I

A. The Conduction Current Equation in the Neutral Ionized Gas

Neglecting the effect of the pressure terms (which seems reasonable, in view of the investigation of the text) we find from (5) of the text

$$\bar{I} = \frac{i\omega}{4\pi} \bar{\alpha} \{ \mathbf{E} - \xi^2(\mathbf{n} \cdot \mathbf{E})\mathbf{n} - i\xi \mathbf{E} \wedge \mathbf{n} \}$$

where

$$\bar{\alpha} = \bar{\sigma}^2 \xi \gamma / (1 - \xi^2); \quad \bar{\sigma} = \bar{\omega}_0/\omega, \quad \xi = \omega_H/(\omega + i\nu) \\ \gamma = \omega/\omega_H.$$

Similarly (using the notation of Part I), we have, for the proton gas

$$\bar{I} = \frac{i\omega}{4\pi} \bar{\alpha} \{ \mathbf{E} - \hat{\xi}^2(\mathbf{n} \cdot \mathbf{E})\mathbf{n} + i\hat{\xi} \mathbf{E} \wedge \mathbf{n} \}$$

where

$$\bar{\alpha} = \bar{\sigma}^2 \hat{\xi} \hat{\gamma} / (1 - \hat{\xi}^2); \quad \bar{\sigma} = \bar{\omega}_0/\omega; \quad \hat{\xi} = \hat{\omega}_H/(\omega + i\hat{\nu}) \\ \hat{\gamma} = \omega/\hat{\omega}_H.$$

Adding these two current vectors to obtain the total current, we find that the equation for \mathbf{E} is now

$$\text{curl curl } \mathbf{E} - \frac{\omega^2}{C^2} \mathbf{E} = - \frac{\omega^2}{C^2} \{ (\bar{\alpha} + \hat{\alpha}) \mathbf{E} \\ - (\bar{\alpha}\xi^2 + \bar{\alpha}\hat{\xi}^2)(\mathbf{n} \cdot \mathbf{E})\mathbf{n} - i(\bar{\alpha}\xi - \bar{\alpha}\hat{\xi}) \mathbf{E} \wedge \mathbf{n} \}.$$

Considering a plane wave with all fields proportional to $\exp(i\mathbf{k}\sqrt{\kappa}\mathbf{z})$, we find from this that

$$\kappa - 1 + \bar{\alpha}_e + \bar{\alpha}_e = -(\bar{\alpha}_e\xi - \bar{\alpha}_e\hat{\xi}).$$

Hence, neglecting ν and $\hat{\nu}$,

$$\kappa = 1 - \frac{\bar{\omega}_e^2}{\omega(\omega_H - \omega)} \left\{ 1 - \frac{m}{\hat{m}} \frac{1-\gamma}{\frac{m}{\hat{m}} + \gamma} \right\}$$

which approximates to the value of κ used in the text when $\omega \gg \hat{\omega}_H$.

APPENDIX II

A. The Coefficients in the Series for \mathbf{F}

Multiplying (41b) by i and subtracting it from (41a)

$$(2n+2)ika\sqrt{\kappa}\nu_n + (2n+2)\{2n+3\}i\mu_{n+1} + \lambda_{n+1} \\ - k^2a^2\hat{\Delta}(1+\xi) \sum_{s=1}^n (\lambda_{n-s} - i\mu_{n-s})f_s = 0.$$

Hence

$$(2n+3)i\mu_{n+1} + \lambda_{n+1} \\ = -ika\sqrt{\kappa}\nu_n + \frac{1}{2n+2}k^2a^2\hat{\Delta}(1+\xi)\sum_{s=1}^n(\lambda_{n-s} - i\mu_{n-s})f_s.$$

Substituting this back in (41a),

$$2nika\sqrt{\kappa}\nu_n + \frac{1}{2n+2}k^2a^2(1+\xi)\sum_{s=1}^n(\lambda_{n-s} - i\mu_{n-s})f_s \\ + k^2a^2\left\{\bar{\alpha}_c\xi(-\lambda_n - i\mu_n) - \bar{\Delta}\sum_{s=1}^n(\lambda_{n-s} - i\mu_{n-s})f_s\right\} = 0.$$

Hence

$$2n\nu_n = -\frac{ika}{\sqrt{\kappa}}\left\{\bar{\alpha}_c\xi(\lambda_n + i\mu_n) \right. \\ \left. + \bar{\Delta}\left(1 - \frac{1+\xi}{2n+2}\right)\sum_{s=1}^n(\lambda_{n-s} - i\mu_{n-s})f_s\right\}.$$

Increasing n to $n+1$ in this and then substituting in (41c),

$$2(n+2)\frac{ika}{\sqrt{\kappa}}\left\{\bar{\alpha}_c\xi(\lambda_{n+1} + i\mu_{n+1}) \right. \\ \left. + \bar{\Delta}\left(1 - \frac{1+\xi}{2n+4}\right)\sum_{s=0}^n(\lambda_{n-s} - i\mu_{n-s})f_{s+1}\right\} \\ + k^2a^2(\bar{\sigma}_c^2\xi\gamma - 1)\nu_n + k^2a^2\xi\gamma\bar{\delta}_c\sum_{s=1}^nf_s\nu_{n-s} \\ + ika\sqrt{\kappa}\{(2n+3)\lambda_{n+1} + i\mu_{n+1}\} = 0.$$

Substituting in this equation for λ_{n+1} from the second equation of this Appendix, and writing

$$C = -ika\sqrt{\kappa}\nu_n + \frac{1}{2n+2}k^2a^2\bar{\Delta}(1+\xi)\sum_{s=1}^n(\lambda_{n-s} - i\mu_{n-s})f_s,$$

this is,

$$2(n+2)\frac{ika}{\sqrt{\kappa}}\left\{\bar{\alpha}_c\xi[C - 2(n+1)i\mu_{n+1}] \right. \\ \left. + \bar{\Delta}\left(1 - \frac{1+\xi}{2n+4}\right)\sum_{s=0}^n(\lambda_{n-s} - i\mu_{n-s})f_{s+1} \right. \\ \left. + k^2a^2(\bar{\sigma}_c^2\xi\gamma - 1)\nu_n + k^2a^2\xi\gamma\bar{\delta}_c\sum_{s=1}^nf_s\nu_{n-s} \right. \\ \left. + ika\sqrt{\kappa}[-4i(n+1)(n+2)\mu_{n+1} + C(2n+3)] = 0.\right.$$

The coefficient of μ_{n+1} is $4(n+1)(n+2)ka(\sqrt{\kappa} + \bar{\alpha}_c\xi/\sqrt{\kappa})$: hence,

$$4(n+1)(n+2)(\kappa + \bar{\alpha}_c\xi)\mu_{n+1} \\ + i\{\kappa(2n+3) + (2n+2)\bar{\alpha}_c\xi\}C \\ + ka\sqrt{\kappa}(\bar{\sigma}_c^2\xi\gamma - 1)\nu_n + ka\sqrt{\kappa}\xi\gamma\bar{\delta}_c\sum_{s=1}^nf_s\nu_{n-s} \\ - 2i(n+2)\bar{\Delta}\left(1 - \frac{1+\xi}{2n+4}\right)\sum_{s=0}^n(\lambda_{n-s} - i\mu_{n-s})f_{s+1} = 0$$

and

$$\kappa + \bar{\alpha}_c\xi = 1 - \bar{\alpha}_c.$$

B. The Induction

Assume now that μ_n, λ_n are polynomials in X of degree $2n-2$, whose highest term is of the form X^{2n-2}/γ^n times a function of G and γ which is bounded as $\gamma \rightarrow 0$. In view of the occurrence of the factor $ka\sqrt{\kappa}$ in the series for F_3 , the correct form to assume for ν_n will be a polynomial in X of degree $2n-1$ whose highest term is of the form¹ $X^{2n-1}/\gamma^{n-1/2}$ times a bounded function of G and γ in $G \geq 0, 0 \leq \gamma \leq 1$.

Then, inspection of the above equation shows that the term in μ_{n+1} involving the highest power of X will be of the form

"a function of G, γ , bounded as $\gamma \rightarrow 0$, times $X^{2n}\gamma^{-n-1}$ "

and this completes the induction for μ_n . The second equation of this Appendix will then complete the induction for λ_n and then (41c) completes it for ν_n .

When X is moderate and $\bar{\Delta}$ is dominant, we have to consider the coefficients from the point of view of polynomials in $\bar{\Delta}$. The correct forms to assume for the highest term in λ_n and μ_n are

$(\bar{\Delta}/\gamma)^n$ times a bounded function of γ in $0 \leq \gamma \leq 1$

and, for ν_n ,

$\bar{\Delta}^n/\gamma^{n-1/2}$ times a bounded function of γ in $0 \leq \gamma \leq 1$.

Inspection of the above equation then shows (remembering that $\bar{\Delta} \propto \bar{\Delta}$) that the term in μ_{n+1} involving the highest power of $\bar{\Delta}$ will be of the form

(bounded function of γ), times $\bar{\Delta}^{n+1}/\gamma^{n+1}$.

which completes the induction for μ_n .

The induction for λ_n is then completed by the second equation of this Appendix and that for ν_n by (41c). (Regarding the last note that $\bar{\delta}_c \propto \bar{\Delta}/\gamma^2$ but the relevant terms are those in λ_{n+1} and μ_{n+1} .)

¹ $ka\sqrt{\kappa} = X\gamma^{1/2}(1-\gamma)^{-1/2}$ times (a bounded function of γ in $0 \leq \gamma \leq 1$), and refer to the known expressions for λ_2, μ_2 and ν_2 .

The Propagation of Electromagnetic Waves in Ionized Gases

Part III—Propagation Along Moving Columns

F. H. NORTHOVER†

I. THE EQUATION FOR THE FIELD VECTOR \mathbf{E}

NEGLECTING the pressural and gravitational terms and the effects of collisions, and writing all vectors proportional to $\exp(i\omega/C\sqrt{\kappa}z - i\omega t)$, the equations for \mathbf{I} and $\bar{\mathbf{I}}$ are

$$-i\omega\left(1 - \frac{U}{C}\sqrt{\kappa}\right)\{-i\omega\mathbf{I} + (\text{div } \mathbf{I})\mathbf{U}\} = -\frac{1}{4\pi}i\omega\omega_0^2\left(\mathbf{E} + \frac{1}{C}\mathbf{U} \wedge \mathbf{H}\right) + i\omega\omega_H\mathbf{I} \wedge \mathbf{n} \quad (1)$$

$$-\omega^2\bar{\mathbf{I}} = -\frac{1}{4\pi}i\omega\bar{\omega}_0^2\mathbf{E} + i\omega\omega_H\bar{\mathbf{I}} \wedge \mathbf{n}. \quad (2)$$

The Maxwell relation is

$$\text{curl curl } \mathbf{E} - \frac{\omega^2}{C^2}\mathbf{E} = \frac{4\pi i\omega}{C^2}(\mathbf{I} + \bar{\mathbf{I}}) \quad (3)$$

the contributions in $\hat{\mathbf{I}}$ and $\bar{\hat{\mathbf{I}}}$ on the right being omitted since we shall be confining ourselves to cases in which the effects of the positively charged gases can be neglected.

In the present problem, \mathbf{U} is along the direction of \mathbf{H}_0 so that $\mathbf{U} = U\mathbf{n}$, hence, the equation for \mathbf{I} becomes

$$-i\omega(1 - p\sqrt{\kappa})\{-i\omega\mathbf{I} + U(\text{div } \mathbf{I})\mathbf{n}\} = -\frac{1}{4\pi}i\omega\omega_0^2\mathbf{E} + i\omega\omega_H\mathbf{I} \wedge \mathbf{n} \quad (4a)$$

where

$$p = U/C \quad (4b)$$

$$\mathbf{E} = \mathbf{E} - \frac{iU}{\omega}\mathbf{n} \wedge \text{curl } \mathbf{E}, \quad (4c)$$

and C is the velocity of light. Eq. (4a) can now be solved for the vector function \mathbf{I} and, after extensive analysis,

$$\mathbf{I} = \frac{\omega_0^2}{4\pi(\Omega^2 - \omega_H^2)}\left\{i\Omega\mathbf{\Gamma} - \frac{iU\omega_H}{\omega}\mathbf{P} + U\left[\text{div}\left(\mathbf{E} - \frac{U\omega_H}{\omega\Omega}\mathbf{P}\right) + \left(\Gamma_1 - \frac{U\omega_H}{\omega\Omega}P_1\right)\frac{N'}{N}\right]\mathbf{n}\right\} \quad (5a)$$

the suffix referring to a radial component and the dash meaning differentiation radially, and where

$$\left. \begin{aligned} \Omega &= \omega(1 - p\sqrt{\kappa}), & k &= \omega/C; \\ p &= U/C; \\ N &= \text{ionic density of moving gas}; \\ \bar{N} &= \text{ionic density of stationary gas}; \\ \omega_0^2 &= 4\pi Ne^2/m; \\ \bar{\omega}_0^2 &= 4\pi \bar{N}e^2/m; \\ \omega_H &= e|H_0|/mC; \\ \alpha &= \omega_0^2/(\Omega^2 - \omega_H^2); & \bar{\alpha} &= \bar{\omega}_0^2/(\omega^2 - \omega_H^2); \\ \mathbf{\Gamma} &= \mathbf{E} - \frac{i\omega_H}{\Omega}\mathbf{E} \wedge \mathbf{n} - \frac{\omega\omega_H^2}{\Omega^3}(\mathbf{n} \cdot \mathbf{E})\mathbf{n}; \\ \mathbf{P} &= \text{curl } \mathbf{E} + \frac{i\Omega}{\omega_H}\mathbf{n} \wedge \text{curl } \mathbf{E}. \end{aligned} \right\} \quad (5b)$$

Putting $U=0$ and using bars, the solution for the current vector in a stationary medium is deduced as a special case; thus

$$\bar{\mathbf{I}} = \frac{i\bar{\omega}_0^2\omega}{4\pi(\omega^2 - \omega_H^2)}\left\{\bar{\mathbf{E}} - \frac{\omega_H^2}{\omega^2}(\mathbf{n} \cdot \bar{\mathbf{E}})\mathbf{n} - \frac{i\omega_H}{\omega}\bar{\mathbf{E}} \wedge \mathbf{n}\right\}. \quad (6)$$

This agrees with a previous result; see equation (5) in Part II. It must be noted that the above method could be used to evaluate $\hat{\mathbf{I}}$ and $\bar{\hat{\mathbf{I}}}$, and thus the effect of the positive ions could be investigated via the full form of (3). This will not be done in the present paper. Use of (3) now gives the required equation for the electric field vector \mathbf{E} ; it is

$$\text{curl curl } \mathbf{E} + k^2\left[\alpha\left\{\frac{\Omega}{\omega}\mathbf{\Gamma} - \frac{U\omega_H}{\omega^2}\mathbf{P} - \frac{iU}{\omega}\left[\text{div}\left(\mathbf{E} - \frac{U\omega_H}{\omega\Omega}\mathbf{P}\right) + \left(\Gamma_1 - \frac{U\omega_H}{\omega\Omega}P_1\right)\frac{N'}{N}\right]\mathbf{n}\right\} + (\bar{\alpha} - 1)\bar{\mathbf{E}} - \bar{\alpha}\frac{\omega_H^2}{\omega^2}(\mathbf{n} \cdot \bar{\mathbf{E}})\mathbf{n} - \frac{i\omega_H}{\omega}\bar{\alpha}\bar{\mathbf{E}} \wedge \mathbf{n}\right] = 0. \quad (7)$$

II. THE FORM OF THE FIELD

Owing to the complexity of this equation as compared with the corresponding equation for the stationary column case, it is not practicable to give such a careful

† Dept. of Mathematics, Carleton University, Ottawa, Can.

examination of the field as was given in Part II. The difficulties are greatest in evaluating the field near the axis, but it is possible to gain an idea of the asymptotic behavior of the field at long distances from the axis. This will now be investigated.

As in Part II, solutions of the form

$$(F_1, F_2, F_3)e^{i\omega/c\sqrt{\kappa}z+i\phi}$$

are considered where the F 's are functions of r alone. An asymptotic solution, as in Part II, of the form

$$F = e^{iK_1kar} \sum_{n=0} a_n^{(1)} r^{-1/2-n} + e^{iK_2kar} \sum_{n=0} a_n^{(2)} r^{-1/2-n} \quad (8)$$

can now be found where K_1 and K_2 are roots of a certain equation in K , which is the eliminant of the following three equations [which correspond to the simpler equations of Part II (15)].

$$\left. \begin{aligned} -K\sqrt{\kappa}C_0 - (\kappa - 1 + \bar{\alpha}_I + s^2\alpha_I)A_0 \\ \quad - i\xi(\bar{\alpha}_I + s\alpha_I)B_0 = 0 \\ \{K^2 + (\kappa - 1 + \bar{\alpha}_I + s^2\alpha_I)\}B_0 + i\xi(\bar{\alpha}_I + s\alpha_I)A_0 = 0 \\ (K^2 + \bar{\sigma}_I^2 + \sigma_I^2 - 1)C_0 \\ \quad - K(\sqrt{\kappa} - p s\alpha_I)A_0 - iKp\xi\alpha_I B_0 = 0 \end{aligned} \right\} \quad (9)$$

where the suffix I denotes value far from the column's axis, the α 's having already been defined at general points by (5b), and

$$\left. \begin{aligned} \sigma_I &= \omega_I/\Omega \\ s &= \Omega/\omega \\ &= 1 - p\sqrt{\kappa} \end{aligned} \right\} \quad (10)$$

so that $1-s$ is the ratio of the column stream velocity to the axial wave propagation velocity.

After considerable algebra and some reduction, the following equation for K is found:

$$\begin{aligned} K^4(1 - \bar{\alpha}_I - s\alpha_I) \\ + K^2\{(\kappa - 1 + \bar{\alpha}_I + s^2\alpha_I)(2 - \bar{\alpha}_I - s\alpha_I - \bar{\sigma}_I^2 - \sigma_I^2) \\ + \xi(\bar{\alpha}_I + s\alpha_I)(\bar{\alpha}_I + \alpha_I)\} \\ + (1 - \bar{\sigma}_I^2 - \sigma_I^2)\{\kappa - 1 + \bar{\alpha}_I + s^2\alpha_I - \xi(\bar{\alpha}_I + s\alpha_I)\} \\ \cdot \{\kappa - 1 + \bar{\alpha}_I + s^2\alpha_I + \xi(\bar{\alpha}_I + s\alpha_I)\} = 0. \end{aligned} \quad (11)$$

The special case of the stationary column is of sufficient importance to warrant special mention; it is obtained from the above by writing $\alpha_I = 0$. Hence, for stationary columns

$$\begin{aligned} K^4(1 - \bar{\alpha}_I) + K^2\{(\kappa - 1 + \alpha_I)(2 - \bar{\alpha}_I - \bar{\sigma}_I^2) + \xi^2\bar{\alpha}_I^2\} \\ + (1 - \bar{\sigma}_I^2)(\kappa - 1 + \bar{\alpha}_I - \bar{\alpha}_I\xi)(\kappa - 1 + \bar{\alpha}_I + \bar{\alpha}_I\xi) = 0. \end{aligned} \quad (12)$$

A. Statement of Principle and Method

Following Part II, (7), the stationary and moving gas distributions are supposedly given by single analytic

distribution functions $\bar{f}(r/\bar{a})$ and $f(r/a)$ in the form

$$\begin{aligned} \bar{N} &= \bar{N}_I + (\bar{N}_c - \bar{N}_I)\bar{f}(r/\bar{a}) \\ N &= N_I + (N_c - N_I)f(r/a) \end{aligned} \quad (13)$$

where the real functions \bar{f} and f decrease steadily from unity on the axis to zero as $r \rightarrow \infty$. The suffix c denotes value on the axis and the suffix I denotes value far from the axis.

The parameters \bar{a} and a can be thought of as "column thickness parameters," or as "smoothing" parameters. As \bar{a} and a tend to infinity the distributions become smoothed out to the homogeneous distributions $N = N_c$, $\bar{N} = \bar{N}_c$, for all finite r .

B. On the Class of Waves Guided by the Columns

A necessary condition for electromagnetic waves to be guided by the column is for the values of K given by (11) to be complex, for this makes the field decrease radially at long distances from the axis like $r^{-1/2} \exp(-Ar)$, where $A > 0$. Sufficient conditions can only be obtained through close consideration of the flux of energy along the distribution, particularly in the axial regions, as noted already in Part II. It does not seem practicable to take this further step in the general case owing to the great complexity of the general problem and to limitations of space.

Consider now the behavior of the solution when a and \bar{a} are made to tend to infinity. The case of the homogeneous media is then reached and, by analysis similar to the preceding, the asymptotic behavior of the vector wave function $F(r)$ far from the axis of the column, must be in the form $r^{-1/2} \exp(iK_1kar)$, where K_1 satisfies an equation similar to (11) but with the suffix I replaced by c ; that is, K_1 satisfies

$$\begin{aligned} K_1^4(1 - \bar{\alpha}_c - s\alpha_c) \\ + K_1^2\{(\kappa - 1 + \bar{\alpha}_c + s^2\alpha_c)(2 - \bar{\alpha}_c - s\alpha_c - \bar{\sigma}_c^2 - \sigma_c^2) \\ + \xi(\bar{\alpha}_c + s\alpha_c)(\bar{\alpha}_c + \alpha_c)\} \\ + (1 - \bar{\sigma}_c^2 - \sigma_c^2)\{\kappa - 1 + \bar{\alpha}_c + s^2\alpha_c - \xi(\bar{\alpha}_c + s\alpha_c)\} \\ \cdot \{\kappa - 1 + \bar{\alpha}_c + s^2\alpha_c + \xi(\bar{\alpha}_c + s\alpha_c)\} = 0. \end{aligned} \quad (14)$$

It seems physically obvious that, if the column is "smoothed out" by making a and \bar{a} tend to infinity, any modes originally guided would eventually be made to "leak" radially. Hence, it seems reasonable to require that the values of K_1 should be real in addition to the requirement that the values of K should be complex. These two conditions impose restrictions upon the coefficients of the two quartics (11) and (14) which, when worked out, give rise to interesting restrictions upon the propagation number κ which, in general, is forced to lie between definite limits which are functions of the frequency. This idea has been worked out by the writer for the simpler case of the stationary column and the results are listed in the Appendix. It is hoped that this may be generalized for the moving column in a subsequent paper.

The present paper, however, shall be confined to a consideration of the case of the propagation of a "standard-type" disturbance; *i.e.*, of a disturbance such that, when a and \bar{a} tend to infinity, it reduces to the axial plane wave of Part I, Section IV, traveling through a homogeneous compound distribution for which $N=N_e$, $\bar{N}=\bar{N}_e$. Eq. (38) of that part gives the value which we must take for κ : the equation is evidently equivalent to

$$\kappa - 1 + \bar{\alpha}_e + s^2\alpha_e + \xi(\bar{\alpha}_e + s\alpha_e) = 0 \quad (15)$$

which, by (14), is consistent with a zero value for K_1 . This is as it should be.

Owing to the complexity of the problem, we shall not immediately begin by investigating the conditions under which the roots of (11) are complex (necessary conditions for guiding). It will be easier (and also of interest) to first investigate the circumstances under which these roots are sensibly the same as those appropriate to the stationary column for the standard type propagation case, as this simpler case has already been fairly fully discussed in Part II. Before doing this, however, the three roots for $\sqrt{\kappa}$ of (36) of Part I, which can be written in the following form, must be considered.

$$x^2 - x_0^2 = \frac{a^2\epsilon_e(1 - px)}{b + px} \quad (16a)$$

where

$$\begin{aligned} a^2 &= \bar{\omega}_e^2/\omega^2, & b &= (\omega_H - \omega)/\omega, \\ \epsilon_e &= \omega_e^2/\bar{\omega}_e^2, & \sqrt{\kappa} &= x, \\ \kappa_0 &= 1 + a^2/b = x_0^2. \end{aligned} \quad (16b)$$

C. The Equation for κ

As far as ionospheric problems and whistling atmospheric are concerned, the most interesting cases are $N \ll \bar{N}$ (and so $\epsilon_e \ll 1$) and $a^2 \gg b$. Of course, $p \ll 1$ always.

If ϵ_e is sufficiently small, the three roots of (16a) approximate to $\pm x_0$ and $-b/p$. The first two represent waves traveling with equal (but opposite) phase velocities equal to the phase velocity of a plane wave in a stationary homogeneous medium with $\bar{N} = \bar{N}_e$, while the root $-b/p$ represents a wave traveling against the ion stream with phase velocity $\gamma/(1-\gamma)$ of the stream velocity $|U|$, where $\gamma = \omega/\omega_H$.

In what follows, we shall confine our attention to propagation against the stream ($U < 0$, $x > 0$). For ϵ_e sufficiently small, there are, then (with a single exception noted below), two possible forward traveling waves given by $x = +x_0$ and $x = b/(-p)$. The exception occurs when the phase velocities (defined by C/x) of these two waves approach equality; true wave propagation in the medium then becomes impossible as the roots x become complex. This can easily be seen from a diagram.

When ϵ_e is sufficiently small, the second approximations to the two roots x under discussion are found to be as follows:

$$\left. \begin{aligned} x &= x_0 \left\{ 1 + \frac{a^2\epsilon_e}{2x_0^2} \frac{1 - px_0}{b + px_0} \right\} \\ &\approx x_0 \left\{ 1 + \frac{b\epsilon_e}{2} \frac{\sqrt{b} - ap}{b^{3/2} + ap} \right\} \\ x &= -\frac{b}{p} \left\{ 1 - \frac{a^2\epsilon_e}{b} \frac{1 + b}{b^2/p^2 - x_0^2} \right\} \\ &\approx -\frac{b}{p} \left\{ 1 - \frac{a^2\epsilon_e(1 + b)p^2}{b^3 - a^2p^2} \right\} \end{aligned} \right\} \quad (17)$$

These approximations obviously fail if $p \rightarrow -b/x_0$; *i.e.*, (effectively) if $p \rightarrow b^{3/2}/a$; *i.e.*, to $(\omega_H - \omega)^{3/2}/\bar{\omega}_e\sqrt{\omega}$. The phase velocities (C/x) of the two forward waves then approach equality and (42) of Part I shows that the two values of x under discussion then become complex.

We begin by investigating the circumstances under which the column type behavior is "quasi-stationary." By this, we shall mean that (11) effectively approximates to (12). It should be noted here that the column actually behaves like a stationary column only for the standard-type propagation waves for which $x = \pm x_0$; this would not be so for the wave $x = -b/p$ even if (11) approximated to (12).

For quasi-stationary type propagation [(11) approximating to (12)], we require that

$$\sigma_I^2/\bar{\sigma}_I^2 \ll 1 \quad (18)$$

and

$$|\alpha_I/\bar{\alpha}_I| \ll 1. \quad (19)$$

The first condition is equivalent to $\epsilon_I/s^2 \ll 1$. This is satisfied under our hypothesis which is $|\epsilon_I| \ll 1$ and $p < 0$, $x > 0$ (so that then $s > 1$).

The second condition is equivalent to

$$\epsilon_I |(1 - \gamma^2)/(1 - s^2\gamma^2)| \ll 1 \quad (20)$$

i.e., since $\epsilon_I \ll 1$, to

$$\epsilon_I |(1 - \gamma)/(1 - s\gamma)| \ll 1$$

i.e., to

$$\epsilon_I b/(b + px) \ll 1.$$

Writing

$$x = -(b/p)(1 - \mu) \quad (21)$$

this condition is equivalent to

$$\epsilon_I/\mu \ll 1. \quad (22)$$

Hence, since ϵ_I is small, propagation along the column will be quasi-stationary except if a particular root x of (16a) were so near to $-b/p$ that μ became comparable with, or small compared with, ϵ_I . We examine this condition further. Eq. (16) may be written

$$x = -\frac{b}{p} \left\{ 1 - \frac{a^2\epsilon_e}{b} \frac{1 - px}{x^2 - x_0^2} \right\} \quad (23)$$

i.e.,

$$\mu = + \frac{a^2 \epsilon_0}{b} \frac{1 - px}{x^2 - x_0^2} \quad (24)$$

$$\doteq + \epsilon_0(1 - px) / \left\{ \left(\frac{x}{x_0} \right)^2 - 1 \right\}.$$

Since in determining our criterion we are concerned (as explained above) only with small values of μ , this gives, effectively,

$$\mu \doteq + \epsilon_0(1 + b) / \left\{ \left(\frac{b}{px_0} \right)^2 - 1 \right\}, \quad (25)$$

provided that the right hand side of this equation is small (e.g., $< \frac{1}{4}$) for otherwise we should be involved in a contradiction in our derivation of (25). What happens if the right of (25) becomes large is that all forward traveling waves disappear, owing to the appropriate roots x of (16) becoming complex.

Hence, subject to the proviso just mentioned, we have

$$\frac{\epsilon_I}{\mu} = \frac{\left(\frac{b}{px_0} \right)^2 - 1}{1 + b} \left(\frac{\epsilon_I}{\epsilon_0} \right)$$

i.e.,

$$\frac{\epsilon_I}{\mu} = \frac{\left(\frac{b}{px_0} \right)^2 - 1}{1 + b} \left(\frac{1 - \Lambda}{1 - \bar{\Lambda}} \right) \quad (26)$$

where, with a similar notation to that of Part II,

$$\Lambda = (N_e - N_I)/N_e, \quad \bar{\Lambda} = (\bar{N}_e - \bar{N}_I)/\bar{N}_e. \quad (27)$$

Propagation, therefore, fails to be quasi-stationary when

$$| (b/px_0)^2 - 1 | / (b + 1) = 0(1) \quad (28)$$

or

$$| (b/px_0)^2 - 1 | / (b + 1) \gg 1. \quad (29)$$

In the latter case, it may be said that the propagation along the column is "anomalous."

Condition (29) for anomalous propagation is easily seen to be (using $a^2 \gg b$) equivalent to

$$| p | \ll (\omega_H - \omega)^{3/2} / \bar{\omega}_e \sqrt{\omega_H}. \quad (30)$$

Columns for which the axial velocity is so slow that (30) is true will be termed "columns with slow drift." Hence, descriptively, for columns with slow drift, there are two waves which are propagated truly quasi-stationarily [i.e., those for which $\sqrt{\kappa} \doteq \pm \bar{\omega}_e / \sqrt{\omega(\omega_H - \omega)}$], but there is in addition a third "slow" wave (anomalous propagation along columns with slow drift) for which

$$\sqrt{\kappa} \doteq - b/p \{ (\omega_H - \omega)/\omega | p | \}. \quad (31)$$

D. Discussion of the Anomalous Wave

For this wave, we must use the condition $\alpha_I \gg \bar{\alpha}_I$ [to which (29) and (30) are equivalent] and the already

established condition $\sigma_I^2 \ll \bar{\sigma}_I^2$ in (11) in order to obtain the exponent numbers K which govern the axial spreading of the field; thus the approximations obtained are

$$K^2 \doteq 4\Lambda \frac{1 - \bar{\Lambda}}{1 - \Lambda} \frac{\omega_H^2 \omega_e^2}{\omega^4 \left\{ (1 + \Lambda) \frac{\omega_H}{\omega} + (1 - \Lambda) \right\}} \quad (32)$$

or

$$\doteq \frac{\omega_e^2}{2\omega(\omega_H - \omega)\epsilon_0} \left(\frac{P_0}{p} \right)^2 \left\{ (1 + \Lambda) \frac{\omega_H}{\omega} + (1 - \Lambda) \right\}$$

where

$$P_0 = (\omega_H - \omega)^{3/2} / \bar{\omega}_e \sqrt{\omega_H} \quad (32a)$$

and used in the working was the fact that $(\Omega - \omega_H)/\omega_H \ll 1$ (equivalent to $\mu \ll 1$); in fact, the result

$$\Omega - \omega_H = -b\omega\mu$$

$$\doteq -\epsilon_0(\omega_H - \omega)(p/P_0)^2 \text{ for this case,} \quad (33)$$

was used.

It is at once apparent that, for radial exponential attenuation of the field (which in general is expressed asymptotically for large r in terms of both of these roots in K^2), we require

$$\Lambda < 0 \quad (34)$$

and this attenuation will only be appreciable (exceed π nepers when $r=a$) if

$$\left. \begin{aligned} 4(-\Lambda)^{1/2} \frac{\omega_H}{\omega} X \left\{ (1 + \Lambda) \frac{\omega_H}{\omega} + (1 - \Lambda) \right\}^{-1/2} &> 1 \\ \left(\frac{2\omega}{\omega_H - \omega} \right)^{1/2} \frac{XP_0}{p} \left\{ (1 + \Lambda) \frac{\omega_H}{\omega} + (1 - \Lambda) \right\}^{1/2} &> 1 \end{aligned} \right\} \quad (35)$$

where $X = a/\lambda_e$, $\omega_e/C = 2\pi/\lambda_e$. For (35) to be satisfied over most of the frequency range $0 < \omega < \omega_H$,

$$\left. \begin{aligned} X(P_0/p) &> 1 \\ X\sqrt{(-\Lambda)} &> 1 \end{aligned} \right\} \quad (36)$$

must hold. It is of interest to notice that

- 1) there does not appear to be a low frequency cutoff for this "anomalous" wave, and
- 2) provided that X is not small compared with unity, large values of $|\Lambda|$ are apparently not needed.

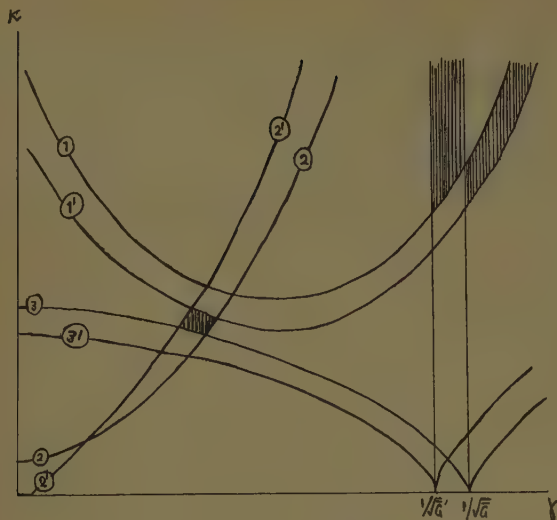
A complete investigation, however, would require the determination of the energy flux distribution through the column, as was attempted in Part II, but, for the general problem considered in Part III, this appears too formidable a task.

SUMMARY

If

$$| (b/px_0)^2 - 1 | / (b + 1) \gg \sqrt{2\epsilon_0}$$

the roots x of (16) approximate closely to $\pm x_0$, $-b/p$.

Fig. 1—Case $\bar{\Lambda} > 0$.

If further,

$$\sqrt{2\epsilon_0} \ll |(b/px_0)^2 - 1| / (b+1) \ll 1$$

then the propagation along the column is truly quasi-stationary (column behaves like a stationary column) for the waves given by $x = \pm x_0$ [(11) then approximating the simpler (12)]. The propagation of the wave for which $x = -b/p$, as regards amount of axial spreading, has not been investigated [one would need to make the approximation $x = -b/p$ in (12)], but the work of Part II shows that the waves $x = \pm x_0$ would be carried along the column (if of central ionic surplus type and sufficiently developed) quasi-stationarily. If

$$|(b/px_0)^2 - 1| / (b+1) = O(1) \\ \text{or } \gg 1$$

propagation is still truly quasi-stationary for the waves for which $x = \pm x_0$ but is not so for the wave $x = -b/p$, as, for this wave, (11) does not approximate (12). The case when $|(b/px_0)^2 - 1| / (b+1)$ is $O(1)$ has not been investigated owing to the complexity of the equations. However, if

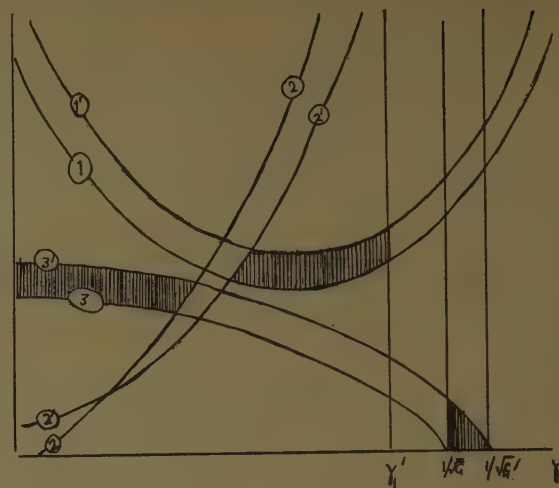
$$|(b/px_0)^2 - 1| / (b+1) \gg 1$$

and if, in addition, $\bar{\Lambda} < 0$, the wave $x = -b/p$ is exponentially attenuated radially and likely to be well guided under relatively light restrictive conditions. This is in contrast to propagation down a stationary column under $x = \pm x_0$, where the most favored columns are those for which $\bar{\Lambda} > 0$.

APPENDIX

ALLOWED VALUES OF κ AND λ FOR THE STATIONARY COLUMN

Key: The following key refers to Figs. 1 and 2.

Fig. 2—Case $\bar{\Lambda} < 0$; $0 < \gamma_1' < 1$ where $1 - 2\gamma_1'^2 - 2\bar{G}'\gamma_1'^2(1 - \gamma_1'^2) = 0$.

$$\frac{1}{\bar{G}\gamma(1 - \gamma)} \quad (1)$$

$$\frac{2(1 - \bar{G}\gamma^2)}{\bar{G}\{1 - 2\gamma^2 - 2\bar{G}\gamma^2(1 - \gamma^2)\}} \quad (2)$$

$$\frac{2}{\bar{G}} \{1 - \bar{G}\gamma^2 - \sqrt{(1 - \bar{G}\gamma^2)\{1 + \bar{G}(1 - \gamma^2)\}}\} \quad (3)$$

where

$$\bar{G} = \omega_H^2 / \bar{\omega}_e^2.$$

The dashed numbers in Figs. 1 and 2 represent graphs of the above functions with \bar{G} replaced by \bar{G}' , where

$$\bar{G}' = \omega_H^2 / \bar{\omega}_T^2 = \bar{G} / (1 - \bar{\Lambda}).$$

ACKNOWLEDGMENT

It is a pleasure to acknowledge here my indebtedness to L. R. O. Storey of the National Bureau of Standards, Boulder, Colo., who suggested the investigation; C. Hines (who suggested the extension of the theory to moving columns); R. Barrington of the Canadian Defence Research Board, Ottawa, Can., for helpful discussions and suggestions; and M. Kline and B. Friedman of the Division of Electromagnetic Research, Institute of Mathematical Sciences, New York University, New York, N. Y., for some useful discussions.

BIBLIOGRAPHY

- [1] R. A. Helliwell, "Low Frequency Propagation Studies—Part I. Whistlers and Related Phenomena," AFCRC, Cambridge, Mass., final rept., TR-189, AD 110184; June 15, 1953 to Sept. 30, 1956.
- [2] L. R. O. Storey, "A Method to Interpret the Dispersion Curves of Whistlers," Radio Phys. Lab., Defence Res. Board Telecommun. Establ., Ottawa, Can., Rept. No. 23-4-1; April, 1958.
- [3] R. L. Smith, "The Guiding of Whistlers," *Proc. Boulder Symp. on the Propagation of VLF Radio Waves*; January 23-26, 1957.

Antennas on Circular Cylinders

H. LOTTRUP KNUDSEN†

Summary—On the basis of the results obtained by Silver and Saunders [4] for the field radiated from an arbitrary slot in a perfectly conducting circular cylinder, expressions have been derived for the field radiated by a narrow helical slot, with an arbitrary aperture field distribution, in a circular cylinder. The cases of a standing wave and a progressing wave aperture field are given particular consideration. It has also been indicated how a finite width of the slot can be taken into consideration. The results for the helical slot have been used for calculating the field radiated from a *U*-shaped slot antenna in a circular cylinder.

By a procedure similar to the one used by Silver and Saunders, expressions have been derived for the field radiated from an arbitrary surface current distribution on a cylinder surface coaxial with a perfectly conducting cylinder. The cases where the space between the two cylindrical surfaces have the same characteristic constants and different constants are treated separately.

Extensive numerical computations of the field radiated from the slot antennas described here are being carried out, but no numerical results are yet available.

INTRODUCTION

DURING the last ten or fifteen years, there has been a great interest in antennas on or near circular cylinders. In particular, a large number of papers have dealt with the radiation from slots in such cylinders. In 1958 the author [1] gave a survey of the results of this work. The report containing this survey was made for the Martin Company, Littleton, Colo., by the P.E.C. Corporation, Boulder, Colo., and published by this corporation. This report contains references to the literature, working formulas for the radiated fields, and a compilation of graphs of radiated fields, many of these graphs published for the first time there. No derivations of the formulas are given; in collecting the material for this report, a bibliography published by Wait [2] in 1957 was found very helpful.

Wait [3] has published a very useful monograph in which a large amount of material is compiled regarding the fields radiated from slots in circular cylinders and other types of cylinders; derivations of the formulas are given in the necessary detail, and the meaning of the formulas are in many cases made clear by graphs. Much of the material in the book is due to his own investigations.

The present paper deals with the field radiated from various types of antennas on or near circular cylinders. The work described is a minor part of some work which the author undertook for the P.E.C. Corporation on a contract with the Martin Company from September, 1958, until June, 1959.

Among the many excellent papers published during the last decade on the field radiated from slot antennas in circular cylinders one paper, by Silver and Saunders

[4], is outstanding in this author's opinion. Their paper has served, or could have served, as a model for any paper dealing with the field radiated from a slot, with a prescribed aperture field in a circular cylinder having a diameter of the same order of magnitude as the wavelength. It has also served as the basis for all our slot antenna computations and as a model for the computations made here regarding the field radiated from wire antennas near circular cylinders. We shall, therefore, start by stating the result of Silver and Saunders in a form suitable for our own use.

ARBITRARY SLOT IN CIRCULAR CYLINDER

For convenient reference, we shall state here the formula derived by Silver and Saunders for the field radiated by an arbitrary slot in an infinitely long circular cylinder. Consider a circular cylinder with radius a as shown in Fig. 1. A rectangular coordinate system

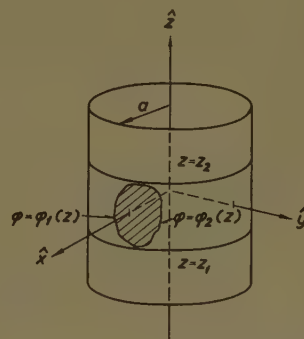


Fig. 1—Arbitrary slot in circular cylinder.

(x, y, z) is introduced with the axis of the cylinder as the z axis, and we further introduce a cylindrical coordinate system (ρ, ϕ, z) and a spherical coordinate system (r, θ, ϕ) related in the conventional way with the rectangular coordinate system. On the surface of the cylinder is a slot of arbitrary shape; the slot is situated in the region between $z = z_1$ and $z = z_2$ ($z_1 < z_2$), and it is bounded by the curves $\phi = \phi_1(z)$ and $\phi = \phi_2(z)$.

It is assumed that the tangential component of the electric field strength in the slot is described by

$$E_\phi = \begin{cases} F_1(\phi, z) & \text{in the aperture,} \\ 0 & \text{outside the aperture,} \end{cases}$$

$$E_z = \begin{cases} F_2(\phi, z) & \text{in the aperture,} \\ 0 & \text{outside the aperture,} \end{cases}$$

where $F_1(\phi, z)$ and $F_2(\phi, z)$ are prescribed functions of ϕ and z .

† The Technical University of Denmark, Copenhagen, Denmark.

Here, and in what follows, we use as a time factor $e^{j\omega t}$ and define the wave number $k=2\pi/\lambda$ where λ is the wavelength. The components of the electric field strength in the far zone field may then be expressed as

$$E_r = 0,$$

$$E_\theta = -\frac{e^{-jk r}}{2\pi^2 r} \sum_{n=-\infty}^{\infty} \frac{j^{n+1} e^{-jn\phi}}{\sin \theta H_n^{(2)'}(ka \sin \theta)} I_n^{(2)},$$

$$E_\phi = \frac{e^{-jk r}}{2\pi^2 r} \sum_{n=-\infty}^{\infty} \frac{j^n e^{-jn\phi}}{H_n^{(2)'}(ka \sin \theta)} \left[I_n^{(1)} + \frac{n \cos \theta}{ka \sin^2 \theta} I_n^{(2)} \right],$$

where

$$I_n^{(\kappa)} = \int_{\xi_1}^{\xi_2} d\xi e^{jk\xi \cos \theta} \int_{\phi_1(\xi)}^{\phi_2(\xi)} F_\kappa(\beta, \xi) e^{jn\beta} d\beta \quad \kappa = 1 \text{ and } 2.$$

In carrying out the above integration, it is convenient to be able to use any orthogonal coordinates (u_1, u_2) on the cylinder surface which fit the geometry of the slot. For this purpose, we redefine the components of the aperture field \bar{F} by

$$E_\phi = \begin{cases} F_1(u_1, u_2) & \text{in the aperture,} \\ 0 & \text{outside the aperture,} \end{cases}$$

$$E_z = \begin{cases} F_2(u_1, u_2) & \text{in the aperture,} \\ 0 & \text{outside the aperture.} \end{cases}$$

An element of length dl on the cylinder surface $r=a$ may now be expressed by

$$dl = \sqrt{(h_1 du_1)^2 + (h_2 du_2)^2},$$

where $h_1 = h_1(u_1, u_2)$ and $h_2 = h_2(u_1, u_2)$ are functions of the coordinates u_1 and u_2 . We further express ξ and β as functions of u_1 and u_2 ,

$$\xi = \xi(u_1, u_2),$$

$$\beta = \beta(u_1, u_2).$$

Often the tangential electric field $\bar{F}(u_1, u_2)$ in the aperture is expressed as a certain constant voltage V_0 divided by a certain constant length w and multiplied by a normalized dimension-free field distribution function $\bar{f}(u_1, u_2)$

$$\bar{F}(u_1, u_2) = \frac{V_0}{w} \bar{f}(u_1, u_2).$$

It may be convenient to further express the far zone field $\bar{E}(r, \theta, \phi)$ in the following way

$$\bar{E}(r, \theta, \phi) = V_0 \frac{e^{-jk r}}{r} \bar{e}(\theta, \phi).$$

The normalized electric field strength $\bar{e}(\theta, \phi)$, which does not depend upon r , has the components

$$e_r = 0,$$

$$e_\theta = -\frac{1}{2\pi^2} \sum_{n=-\infty}^{\infty} \frac{j^{n+1} e^{-jn\phi}}{\sin \theta H_n^{(2)'}(ka \sin \theta)} i_n^{(2)},$$

$$e_\phi = \frac{1}{2\pi^2} \sum_{n=-\infty}^{\infty} \frac{j^n e^{-jn\phi}}{H_n^{(2)'}(ka \sin \theta)} \left[i_n^{(1)} + \frac{n \cos \theta}{ka \sin^2 \theta} i_n^{(2)} \right],$$

where $i_n^{(\kappa)}$ are dimension-free quantities given by

$$i_n^{(\kappa)} = \frac{1}{aw} \iint f_\kappa(u_1, u_2) \cdot e^{j[k\xi(u_1, u_2) \cos \theta + n\beta(u_1, u_2)]} h_1(u_1, u_2) h_2(u_1, u_2) du_1, du_2$$

$\kappa = 1 \text{ and } 2.$

This is a convenient formulation of Silver and Saunders' results for many practical purposes.

HELICAL SLOT IN CIRCULAR CYLINDER

Arbitrary Aperture Field

Although there is a considerable practical interest attached to the use of inclined slots in a circular cylinder, to the author's knowledge no analytical expression for, and no numerical computation of, the field radiated from an inclined slot in a circular cylinder have been published so far. The content of this section is a revised edition of part of a Master's thesis by Pedersen [5]; the thesis work was done under the author's direction in 1958.

With reference to Fig. 2, let us first consider an in-

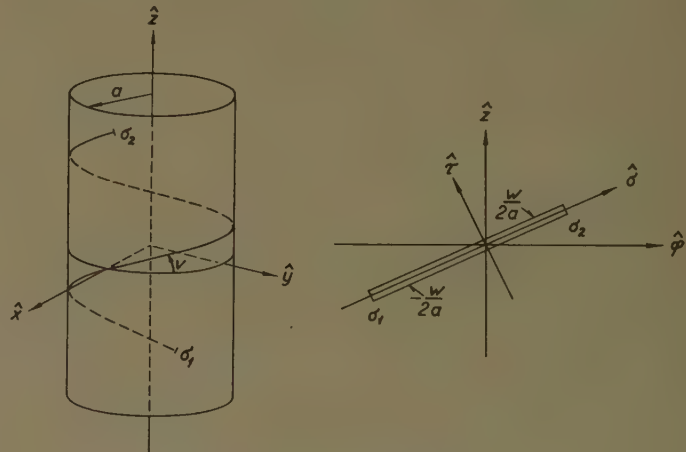


Fig. 2—Helical slot in circular cylinder.

finitely narrow, helical slot in an infinitely long, circular cylinder, and let us assume that the field in the slot is prescribed. The radius of the cylinder is called a . The centerline of the slot is assumed to be a piece of a helix with its midpoint passing through the point $(x, y, z) = (a, 0, 0)$ as shown in Fig. 2. The angle which the helix forms with a plane transverse to the axis is called v . We introduce a coordinate σ along the slot so that $s=a\sigma$ is the length of arc in the σ direction. The ends of the slot are assumed to be situated at

$$\sigma = \pm \sigma_1 = \pm \frac{s_1}{2a}.$$

The tangential component of the aperture field is assumed to be transverse to the slot. Denoting by $\hat{\tau}$ a unit vector pointing in such a direction transverse to the slot that $\hat{\sigma} \times \hat{\tau}$ is a unit vector pointing normally outwards from the cylinder surface, we may express the tangential aperture field by

$$\bar{F} = \frac{V_0 f(\sigma)}{w} \hat{\tau},$$

where V_0 has the dimension of voltage and w the dimension of length, whereby $f(\sigma)$ becomes dimension-free. We may think of w as the width of the slot and of V_0 as a reference voltage across the slot.

Using the general formulas for the field radiated from an arbitrary slot in a circular cylinder given in the last section, we may now obtain the following expressions for the components of the normalized electric field strength radiated from the inclined slot

$$e_\theta = -\frac{\cos v}{2\pi^2 \sin \theta} \sum_{n=-\infty}^{\infty} \frac{j^{n+1} e^{-jn\phi} j_n}{H_n^{(2)}(ka \sin \theta)},$$

$$e_\phi = \frac{1}{2\pi^2} \sum_{n=-\infty}^{\infty} \frac{j^n e^{-jn\phi} j_n}{H_n^{(2)'}(ka \sin \theta)} \left[-\sin v + \frac{n \cos v \cos \theta}{ka \sin^2 \theta} \right],$$

where

$$j_n = \int_{-\sigma_1}^{\sigma_1} f(\sigma) e^{iK_n \sigma} d\sigma$$

with

$$K_n = ka \sin v \cos \theta + n \cos v.$$

When the aperture field is given, *i.e.*, when $f(\sigma)$ is known, the integral j_n may be computed and the field found from the above formulas.

So far, we have considered only infinitely narrow slots. It is desirable also to have formulas available for the field radiated from an inclined slot of finite width. Considering an inclined slot with a finite width w , as shown in Fig. 3, and assuming that the tangential aper-

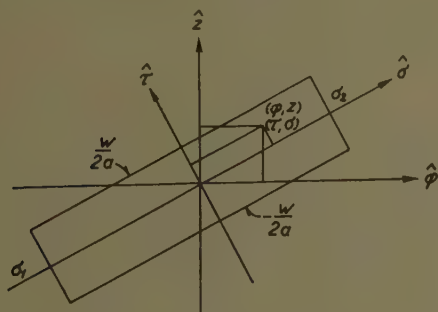


Fig. 3—Inclined slot with finite width.

ture field is transverse to the longitudinal direction of the slot and that the field does not vary in the transverse direction, we find that the formulas for the radiated fields given above should be replaced by

$$e_\theta = -\frac{\cos v}{2\pi^2 \sin \theta} \sum_{n=-\infty}^{\infty} \frac{j^{n+1} e^{-jn\phi} j_n}{H_n^{(2)}(ka \sin \theta)} P_n,$$

$$e_\phi = \frac{1}{2\pi^2} \sum_{n=-\infty}^{\infty} \frac{j^n e^{-jn\phi} j_n}{H_n^{(2)'}(ka \sin \theta)} \left[-\sin v + \frac{n \cos v \cos \theta}{ka \sin^2 \theta} \right] P_n,$$

where

$$P_n = \frac{\sin \frac{H_n w}{2a}}{\frac{H_n w}{2a}}$$

with

$$H_n = ka \cos v \cos \theta - n \sin v.$$

The present formulas differ from the corresponding expressions for an infinitely narrow, inclined slot with the same field distribution only by the factor P_n , which is present here, but absent in the formulas for the infinitely narrow slot. If we let the width, w , of the slot tend towards zero, we find

$$\lim_{w \rightarrow 0} P_n = 1.$$

The expressions for the components e_θ and e_ϕ , derived here, then converge towards the expressions for e_θ and e_ϕ for an infinitely narrow, inclined slot, as they should.

In the following, only the expressions for an infinitely narrow slot will be stated.

Sinusoidal Aperture Field

One use of the above expressions is for obtaining an approximate expression for the field radiated from a plane, inclined slot with a sinusoidal aperture field.

A plane slot in a circular cylinder actually is part of an ellipse. However, when the slot is short, as compared to the circumference of the cylinder, it may be well approximated by part of a helix. The theory given in the last section may, therefore, be used in finding the field radiated from such a slot. The length of the slot is called l , and the center of the slot is assumed to be situated at $(x, y, z) = (a, 0, 0)$. A sinusoidal field distribution in the slot will then be expressed by the normalized field distribution function

$$f(\sigma) = \cos \frac{\pi a}{l} \sigma.$$

Inserting this function in the above formula for the integral j_n , we obtain

$$j_n = \frac{\frac{2\pi a}{l} \cos K_n \frac{l}{2a}}{\left(\frac{\pi a}{l}\right)^2 - K_n^2},$$

where K_n is given above. This expression for j_n should be inserted in the above expressions for the θ - and ϕ -components of the normalized electric field strength \bar{e} .

Setting $v=0$ in the formulas described above, we obtain the well-known formulas for the field radiated from a circumferential slot. On the other hand, setting $v=90^\circ$ in the above mentioned formulas, we obtain expressions which are identical with the well-known formulas for the axial slot found in the literature, except for the fact that the two formulas have opposite signs. This difference in sign is due to the circumstance that here the positive direction for the field in the aperture is chosen as the direction of \hat{s} (which in the case of the axial slots becomes $\hat{s} = -\hat{\phi}$) whereas in most papers dealing with the axial slot, it is customary to choose the direction of $\hat{\phi}$ as the positive direction for the aperture field.

In the above investigation, we have assumed the inclined slot to be helical instead of being an ellipse as it would be if the slot were formed by cutting the cylinder with a plane. The distance Δ between the end points of the actual elliptical slot and the helical slot (see Fig. 4) is approximately

$$\Delta = a \sin v (\phi_m - \sin \phi_m),$$

where

$$\phi_m = \frac{l \cos v}{2a}.$$

For the special case of a half wavelength slot, the relative distance ϵ between the helix and the ellipse at their end points, defined as the distance Δ divided by the length l of the slot, has been plotted in Fig. 5. It is seen from this figure that when the diameter of the cylinder is one wavelength or more, the maximum relative deviation ϵ between the actual and the approximating slot is only a few tenths of one per cent.

From the above expressions for the components of the field radiated by an inclined slot with a sinusoidal aperture field distribution, we may obtain the following symmetry relations

$$e_\theta(\pi - \theta, -\phi) = e_\theta(\theta, \phi),$$

$$e_\phi(\pi - \theta, -\phi) = e_\phi(\theta, \phi).$$

Since these symmetry relations could also have been established directly by inspection of the geometrical properties of the slot and its excitation, the derivation of these relations furnishes a partial check of the above expressions. The preceding symmetry relations seem generally to exhaust the symmetry properties of the field radiated from an inclined slot with a symmetric excitation. However, in special cases further symmetry relations may exist. For a circumferential slot ($v=0$), for example, we have further

$$e_\theta(\theta, -\phi) = e_\theta(\theta, \phi),$$

$$e_\phi(\theta, -\phi) = -e_\phi(\theta, \phi).$$

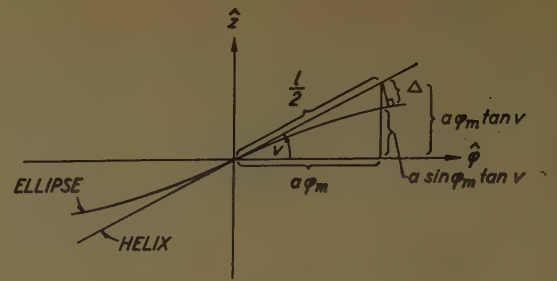


Fig. 4—Relative deviation of helical slot from plane, elliptical slot.

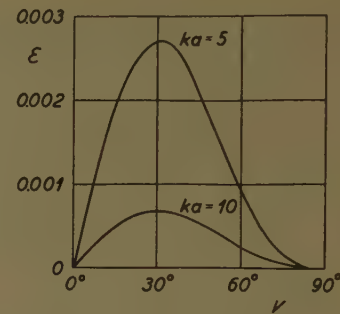


Fig. 5—Relative deviation of helical slot from plane, elliptical slot.

For an axial slot ($v=90^\circ$), we have the following additional symmetry relation

$$e_\phi(\theta, -\phi) = e_\phi(\theta, \phi).$$

All of the previous symmetry relations also hold for the approximate field expressions obtained from the exact expressions by replacing the summation from $-\infty$ to ∞ by a summation from $-N$ to N . The above symmetry relations may therefore be used as a partial check in carrying out the numerical computations.

Progressing Aperture Wave Field

It has been suggested that the field radiated from a helical slot in a circular cylinder with a progressing wave in the aperture be investigated. We assume that the normalized aperture field distribution $f(\sigma)$ is a progressing wave given by

$$f(\sigma) = e^{-ipka\sigma},$$

where p is the quotient between the velocity of light and the velocity of the aperture wave field. Considering only a helical slot with an integral number, N , of turns, we obtain, by using the general expression for the radiated field given above, the following expression for the normalized field $\bar{e}(\theta, \phi)$ radiated from the slot

$$\bar{e}(\theta, \phi) = G(\theta) \bar{e}^*(\theta, \phi - (N-1)\pi),$$

where $G(\theta)$ is the array characteristic of an array of N single turns, and where $\bar{e}^*(\theta, \phi)$ is the field radiated from

a single turn of such a helical slot. The components of the normalized electric field strength $\bar{e}^*(\theta, \phi)$ radiated from a single turn are given by

$$e_\theta^*(\theta, \phi) = -\frac{\sin \pi \mu}{\pi^2 \sin \theta} \sum_{n=-\infty}^{\infty} \frac{e^{jn[(3\pi/2)-\phi]+\pi/2]}{(\mu+n)H_n^{(2)}(ka \sin \theta)},$$

$$e_\phi^*(\theta, \phi) = \frac{\sin \pi \mu}{\pi^2 \cos \nu} \sum_{n=-\infty}^{\infty} \frac{e^{jn[(3\pi/2)-\phi]}}{(\mu+n)H_n^{(2)'}(ka \sin \theta)}$$

$$\cdot \left[-\sin \nu + \frac{n \cos \nu \cos \theta}{ka \sin^2 \theta} \right],$$

where $\mu = ka(\tan \nu \cos \theta - p/\cos \nu)$.

It should be noted that the above relation between the field \bar{e} from a helical slot with N turns and the field \bar{e}^* from a single turn of this slot connects $\bar{e}(\theta, \phi)$, not with $\bar{e}^*(\theta, \phi)$, but with $\bar{e}^*(\theta, \phi - (N-1)\pi)$. This is due to the fact that helical slots with an even number of turns are oriented differently in the coordinate system (x, y, z) from the helical slot with an odd number of turns, in that the end-points of a helical slot with an even number of turns are situated on the line $x=a, y=0$, whereas the end-points of a helical slot with an odd number of turns are situated on the line $x=-a, y=0$. This is illustrated in Fig. 6 for the cases of $N=1$ and $N=2$.

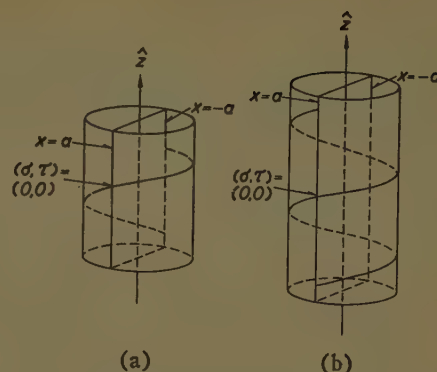


Fig. 6—Helical slot with (a) an odd and (b) an even number of turns.

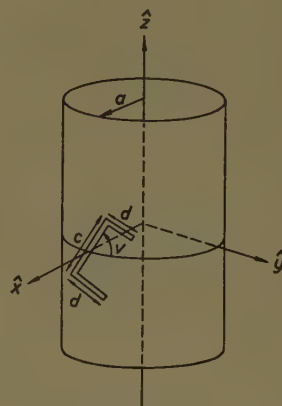


Fig. 7—U-shaped slot in circular cylinder.

U-SHAPED SLOT IN CIRCULAR CYLINDER

One type of antenna that seems to be generally used as a flush-mounted, radiating element in a circular cylinder is a U-shaped slot. In this section, expressions shall be derived for such an antenna under the assumption that the slot is infinitely narrow, and that it is made up of parts of three helical slots. Only the case of a symmetrical slot shall be considered here.

In Fig. 7 is shown a symmetrical, U-shaped slot on the surface of a circular cylinder with radius a . The length of the central part of the slot is called c and the length of the outer parts of the slot d . A coordinate system (x, y, z) is introduced as shown in the figure, so that the midpoint of the slot is located at $(x, y, z) = (a, 0, 0)$. The slot is given a positive direction, and the angle which the central part of the slot so oriented forms with a plane transverse to the axis of the cylinder is called ν . We introduce a dimension-free coordinate σ along the slot with $\sigma=0$ corresponding to one end-point of the slot, and so defined that $s=a\sigma$ is the length of the arc along the slot. The aperture field in the slot is assumed to be everywhere transverse to the longitudinal direction of the slot. The component $F_r(\sigma)$ of the aperture field in the transverse direction, counted positive from the "inside" of the slot towards the "outside," is as-

sumed to be

$$F_r(\sigma) = \frac{V_0}{w} \sin q\sigma,$$

where

$$q = \frac{\pi a}{c + 2d}.$$

By using the general formulas given above for the field radiated from an arbitrary slot with an arbitrary aperture field distribution, we then obtain the following expressions for the components of the normalized electric field strength of the field radiated from the U-shaped slot with a sinusoidal aperture field distribution,

$$e_\theta = -\frac{1}{2\pi^2 \sin \theta} \sum_{n=-\infty}^{\infty} \frac{j^{n+1} e^{-jn\phi}}{H_n^{(2)}(ka \sin \theta)} B_n,$$

$$e_\phi = \frac{1}{2\pi^2} \sum_{n=-\infty}^{\infty} \frac{j^n e^{-jn\phi}}{H_n^{(2)'}(ka \sin \theta)} \left[A_n + \frac{n \cos \theta}{ka \sin^2 \theta} B_n \right],$$

where

$$A_n = j_n^* \cos \nu - j_n^{(2)} \sin \nu,$$

$$B_n = j_n^* \sin \nu + j_n^{(2)} \cos \nu,$$

with

$$j_n^* = \frac{2 \sin \left(K_n^{(2)} \frac{c}{2a} \right)}{q^2 - (K_n^{(1)})^2} \left\{ q \sin \left(K_n^{(1)} \frac{d}{a} \right) - K_n^{(1)} \cos \left(q \frac{c}{2a} \right) + jq \left[\cos \left(K_n^{(1)} \frac{d}{a} \right) - \sin \left(q \frac{c}{2a} \right) \right] \right\},$$

$$j_n^{(2)} = \frac{2}{q^2 - (K_n^{(2)})^2} \left\{ q \cos \left(K_n^{(2)} \frac{c}{2a} \right) \sin \left(q \frac{c}{2a} \right) - K_n^{(2)} \sin \left(K_n^{(2)} \frac{c}{2a} \right) \cos \left(q \frac{c}{2a} \right) \right\},$$

$$K_n^{(1)} = ka \cos v \sin \theta - n \sin v,$$

$$K_n^{(2)} = ka \sin v \cos \theta + n \cos v.$$

A partial check of the above formulas may be obtained by applying them to the case of $d=0$, in which case the U-shaped slot degenerates into an ordinary inclined slot with the length c . It may readily be verified that in this case the above formulas simplify to the formulas obtained previously for a simple inclined slot.

ARBITRARY CURRENT DISTRIBUTION ON CIRCULAR CYLINDRICAL SURFACE COAXIAL WITH CIRCULAR, CONDUCTING CYLINDER

The author has been asked to solve the problem of computing the field radiated from a helical wire, with a progressing current wave, coaxial with a conducting, circular cylinder. This problem may be considered a particular case of the more general problem of finding the field radiated from an arbitrary, given distribution of surface current on a circular cylinder coaxial with the conducting, circular cylinder. Since the latter, more general problem may be solved with only a little more work than required for solving the particular problem of a helical current distribution, and since the solution of the general problem may also have practical applications to other problems, we shall here solve this problem.

Consider a perfectly conducting cylinder with radius a and a coaxial, circular cylindrical surface with radius b as shown in Fig. 8. In formulating the problem, we shall use cylindrical coordinates (ρ, ϕ, z) with the axis of the cylinder as the z axis, whereas for expressing the far zone field, we shall use spherical coordinates (r, θ, ϕ) , these coordinates being related to the cylindrical coordinates in the usual way. On the cylindrical surface with radius b , a distribution of electric current is assumed to exist. The surface current density is assumed to be equal to zero outside a finite region (the aperture), the circumference of which is given by the curves $\phi = \phi_1(z)$ and $\phi = \phi_2(z)$, these curves being confined to

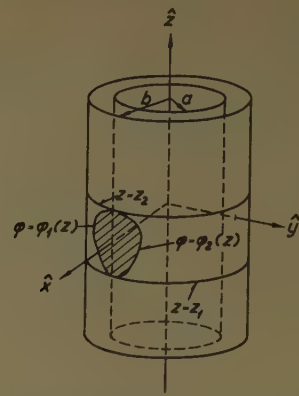


Fig. 8—Surface current distribution on circular cylindrical surface coaxial with conducting, circular cylinder.

the region $z_1 \leq z \leq z_2$. The surface current density $\bar{K}(\phi, z) = K_\phi(\phi, z)\hat{\phi} + K_z(\phi, z)\hat{z}$ is given by

$$K_\phi = \begin{cases} G_1(\theta, \phi) & \text{in the aperture,} \\ 0 & \text{outside the aperture,} \end{cases}$$

$$K_z = \begin{cases} G_2(\theta, \phi) & \text{in the aperture,} \\ 0 & \text{outside the aperture,} \end{cases}$$

where $G_1(\theta, \phi)$ and $G_2(\theta, \phi)$ are prescribed functions. The problem is to find the field that satisfies Maxwell's equations in the regions $a < \rho < b$ and $b < \rho < \infty$, the boundary conditions at the metallic surface $\rho = a$, the saltus conditions at the surface $\rho = b$, and the radiation conditions at infinity (*i.e.*, for $r \rightarrow \infty$). The boundary condition at $r = a$ is expressed by

$$\left. \begin{matrix} E_\phi = 0 \\ E_z = 0 \end{matrix} \right\} \quad \text{for } \rho = a,$$

and the saltus condition at $\rho = b$ is expressed by

$$\left. \begin{matrix} -H_{1\phi} + H_{2\phi} = K_z \\ H_{1z} - H_{2z} = K_\phi \\ E_{1\phi} - E_{2\phi} = 0 \\ E_{1z} - E_{2z} = 0 \end{matrix} \right\} \quad \text{for } \rho = b,$$

where index 1 refers to the inside and index 2 to the outside of the surface $r = b$.

If the field set up by a Hertz dipole near a conducting circular cylinder were known, the solution to the problem formulated above could be obtained as a vectorial superposition of the fields set up by the infinitely many dipoles, into which the present current distribution may be decomposed. Now, the far zone field of a Hertz dipole near a conducting circular cylinder has been obtained by Carter [6], who first computed the current induced in a Hertz dipole near a cylinder by an incident plane wave and then used the reciprocity theorem. However,

for the following three reasons, we shall here use a more direct approach, without making use of Carter's results.

- 1) It is inherent in Carter's derivation that his results can be used only for finding the far zone field. For some problems, it may be desirable to have formulas available for the field at any distance from the current distribution.
- 2) A certain amount of work is necessary for arriving at the desired result using Carter's formulas. It seems, therefore, more motivated to spend the energy in making a direct derivation of the formulas instead of using a method which depends upon the use of the reciprocity theorem.
- 3) As has been pointed out elsewhere by the author [7] and by Nielsen [8], Carter's paper, the excellence of which is undisputed, contains many trivial errors. It seems, therefore, preferable to use a different and more direct method for obtaining the desired formulas.

The method which will be used for solving the present problem is completely analogous to the method used by Silver and Saunders [4] in finding the field radiated from an arbitrary slot in a circular cylinder. The surface current density on the cylinder may be expressed by the following expansion

$$K_{\kappa}(\phi, z) = \frac{1}{4\pi^2} \sum_{n=-\infty}^{\infty} e^{-in\phi} \int_{-\infty}^{\infty} dh \int_{z_1}^{z_2} d\xi \int_{\phi_1(\xi)}^{\phi_2(\xi)} G_{\kappa}(\beta, \xi) e^{in\beta} e^{-ih(z-\xi)} d\beta.$$

The field outside the metallic cylinder is expressed by superposition of basic sets of cylindrical waves (see for example Stratton [9]). We have to use different expansions in the region 1 (*i.e.*, for $a < \rho < b$) and in the region 2 (*i.e.*, for $b < \rho < \infty$). The field in the region α ($\alpha=1$ or 2) may be expressed by

$$E_{\nu}^{(\alpha)}(\rho, \phi, z) = \int_{-\infty}^{\infty} \mathcal{E}_{\nu}^{(\alpha)}(\rho, \phi, z; h) dh,$$

$$H_{\nu}^{(\alpha)}(\rho, \phi, z) = \int_{-\infty}^{\infty} \mathcal{H}_{\nu}^{(\alpha)}(\rho, \phi, z; h) dh,$$

where

$$\alpha = 1 \text{ or } 2 \quad \text{and} \quad \nu = \rho, \phi \text{ or } z.$$

Region 1

$$\mathcal{E}_{\rho}^{(1)} = \left\{ \sum_{n=-\infty}^{\infty} \left[-jh\gamma(a_n J_n'(\gamma\rho) + b_n H_n^{(2)'}(\gamma\rho)) - \frac{n\omega\mu}{\rho}(c_n J_n(\gamma\rho) + d_n H_n^{(2)}(\gamma\rho)) \right] e^{-in\phi} \right\} e^{-ihz}$$

and corresponding expressions for $\mathcal{E}_{\phi}^{(1)}$, $\mathcal{E}_z^{(1)}$, $\mathcal{H}_{\rho}^{(1)}$, $\mathcal{H}_{\phi}^{(1)}$ and $\mathcal{H}_z^{(1)}$,

Region 2

$$\mathcal{E}_{\rho}^{(2)} = \sum_{n=-\infty}^{\infty} \left[-jh\gamma e_n H_n^{(2)'}(\gamma\rho) - \frac{n\omega\mu}{\rho} f_n H_n^{(2)}(\gamma\rho) \right] e^{-in\phi} \left\} e^{-ihz},$$

and corresponding expressions for $\mathcal{E}_{\phi}^{(2)}$, $\mathcal{E}_z^{(2)}$, $\mathcal{H}_{\rho}^{(2)}$, $\mathcal{H}_{\phi}^{(2)}$ and $\mathcal{H}_z^{(2)}$, where

$$\gamma = \sqrt{k^2 - h^2},$$

and where the quantities a_n , b_n , c_n , d_n , e_n , and f_n are constants to be determined from the boundary conditions and the saltus conditions of the problem.

Inserting the previously determined expressions for the field components in the equations expressing the boundary conditions and the saltus conditions, we find

$$\begin{pmatrix} A_n & B_n & C_n & D_n & 0 & 0 \\ E_n & F_n & 0 & 0 & 0 & 0 \\ G_n & H_n & J_n & K_n & L_n & M_n \\ N_n & O_n & 0 & 0 & P_n & 0 \\ Q_n & R_n & S_n & T_n & U_n & V_n \\ 0 & 0 & X_n & Y_n & 0 & Z_n \end{pmatrix} \begin{pmatrix} a_n \\ b_n \\ c_n \\ d_n \\ e_n \\ f_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \Phi_2 \\ \Phi_1 \end{pmatrix},$$

where

$$A_n = -\frac{nh}{a} J_n(\gamma a),$$

$$\dots$$

$$\dots$$

$$R_n = j \frac{k^2 \gamma}{\omega \mu} H_n^{(2)'}(\gamma b),$$

$$\dots$$

$$\dots$$

$$Z_n = -\gamma^2 H_n^{(2)}(\gamma b),$$

and where

$$\Phi_{\kappa} = \frac{1}{4\pi^2} \int_{z_1}^{z_2} d\xi \int_{\phi_1(\xi)}^{\phi_2(\xi)} G_{\kappa}(\beta, \xi) e^{i(n\beta + h\xi)} d\beta \quad \kappa = 1 \text{ or } 2.$$

Solving the preceding system of six linear equations and inserting the expressions for the coefficients a_n to f_n so obtained in the expressions for the field components given before, we obtain the expressions for the components of the electric and the magnetic field strength in regions 1 and 2. As an example of the expressions for these twelve quantities, we here give the expression for $E_{\rho}^{(2)}$,

$$E_p^{(2)} = \frac{j\pi}{2} \int_{-\infty}^{\infty} \left\{ \sum_{n=-\infty}^{\infty} \left[\frac{h(nh\Phi_1 - \gamma^2 b\Phi_2)(-J_n(\gamma b)H_n^{(2)}(\gamma a) + J_n(\gamma a)H_n^{(2)}(\gamma b))H_n^{(2)'}(\gamma \rho)}{k\gamma H_n^{(2)}(\gamma a)} \right. \right. \\ \left. \left. + \frac{nkb\Phi_1(-J_n'(\gamma b)H_n^{(2)'}(\gamma a) + J_n'(\gamma a)H_n^{(2)'}(\gamma b))H_n^{(2)}(\gamma \rho)}{\gamma \rho H_n^{(2)'}(\gamma a)} \right] e^{-in\phi} \right\} e^{-ihz} dh,$$

where $\zeta = \sqrt{\mu/\epsilon}$ is the intrinsic impedance of free space. In the far zone field, those parts of the expressions for the field components in which ρ occurs in the denominator are insignificant as compared to the remaining parts of these expressions. Making use of this fact, we notice that each component of the electric field strength may be expressed as

$$E_\nu(\rho, \phi, z) = \int_{-\infty}^{\infty} \left\{ \sum_{n=-\infty}^{\infty} [s_\nu^* s_{\nu n} H_n^{(2)}(\gamma \rho) + t_\nu^* t_{\nu n} H_n^{(2)'}(\gamma \rho)] e^{-in\phi} \right\} e^{-ihz} dh,$$

where

$$s_{\nu n} = s_{\nu n}^{(1)}\Phi_1 + s_{\nu n}^{(2)}\Phi_2, \\ t_{\nu n} = t_{\nu n}^{(1)}\Phi_1 + t_{\nu n}^{(2)}\Phi_2,$$

and where the index ν stands for one of the cylindrical coordinates ρ, ϕ and z . The symbols $s_{\nu n}$ and $t_{\nu n}$ stand for one of the coefficients e_n and f_n , and the coefficients s_ν^* and t_ν^* are determined from the expressions for the various components.

In the expression for the components in the far zone field, we now approximate the Hankel function and its derivative, both of argument $\gamma\rho$, with the first term of the asymptotic expansions. Introducing further spherical coordinates instead of cylindrical coordinates, we find

$$E_\nu(\rho, \phi, z) = \frac{1}{4\pi^2} e^{j(\pi/4)} \sqrt{\frac{2}{\pi r \sin \theta}} \\ \cdot \sum_{n=-\infty}^{\infty} j^n e^{-jn\phi} \left\{ \int_{z_1}^{z_2} d\xi \int_{\phi_1(\xi)}^{\phi_2(\xi)} G_1(\beta, \xi) e^{jn\beta} d\beta \int_{-\infty}^{\infty} u_1(h) \right. \\ \cdot e^{-j\gamma r(\gamma \sin \theta + h \cos \theta)} dh \\ \left. + \int_{z_1}^{z_2} d\xi \int_{\phi_1(\xi)}^{\phi_2(\xi)} G_2(\beta, \xi) e^{jn\beta} d\beta \int_{-\infty}^{\infty} u_2(h) \right. \\ \cdot e^{-j\gamma r(\gamma \sin \theta + h \cos \theta)} dh \left. \right\},$$

where

$$u_\kappa(h) = [s_\nu^* s_{\nu n}^{(\kappa)} - j t_\nu^* t_{\nu n}^{(\kappa)}] (k^2 - h^2)^{-1/4} e^{jh\xi} \quad \kappa = 1 \text{ or } 2.$$

For large values of r , i.e., for points in the far zone field, approximate expressions for the above integrals with h

as the variable may be obtained by using the saddle-point method. Silver and Saunders [4], in their paper on the field radiated by an arbitrary slot in a circular cylinder, have made a saddle-point evaluation of integrals of the above type. For large values of r , they hereby find

$$\int_{-\infty}^{\infty} u(h) e^{-j\gamma r(\gamma \sin \theta + h \cos \theta)} dh \\ \cong \sqrt{\frac{2\pi k}{r}} \sin \theta u(k \cos \theta) e^{-ikh r} e^{j(\pi/4)}.$$

Using this approximation in the previous expression for E_ν , we obtain the desired far zone expressions for the cylindrical components of the electric field strength. From these formulas, we finally obtain the following expressions for the spherical components of the electric field strength in the far zone field.

$$E_r = 0,$$

$$E_\theta = \frac{e^{-jkr}}{r} \frac{kb}{4\pi} \sum_{n=-\infty}^{\infty} \frac{j^{n+1} e^{-jn\phi} S_n(ka \sin \theta, kb \sin \theta)}{H_n^{(2)}(ka \sin \theta)} \\ \cdot \left\{ \frac{n \cos \theta}{kb \sin \theta} I_n^{(1)} - \sin \theta I_n^{(2)} \right\},$$

$$E_\phi = \frac{e^{-jkr}}{r} \frac{kb}{4\pi} \sum_{n=-\infty}^{\infty} \frac{j^n e^{-jn\phi} V_n(ka \sin \theta, kb \sin \theta)}{H_n^{(2)'}(ka \sin \theta)} I_n^{(1)},$$

where

$$I_n^{(x)} = \int_{z_1}^{z_2} d\xi e^{jh\xi \cos \theta} \int_{\phi_1(\xi)}^{\phi_2(\xi)} G_x(\beta, \xi) e^{jn\beta} d\beta \quad x = 1 \text{ or } 2,$$

and where we have introduced the functions

$$S_n(x, y) = J_n(x)H_n^{(2)}(y) - H_n^{(2)}(x)J_n(y),$$

$$V_n(x, y) = J_n'(x)H_n^{(2)'}(y) - H_n^{(2)'}(x)J_n'(y).$$

In carrying out this integration, it is convenient to be able to use any orthogonal coordinates (u_1, u_2) on the cylinder surface which fit the geometry of the aperture. For this purpose, we redefine the components G_κ of the surface current distribution \vec{G} on the cylindrical surface $r=b$ by

$$K_\phi = \begin{cases} G_1(u_1, u_2) & \text{in the "aperture,"} \\ 0 & \text{outside the "aperture,"} \end{cases}$$

$$K_z = \begin{cases} G_2(u_1, u_2) & \text{in the "aperture,"} \\ 0 & \text{outside the "aperture."} \end{cases}$$

An element of length dl on the cylinder surface $r=b$ may now be expressed by

$$dl = \sqrt{(h_1 du_1)^2 + (h_2 du_2)^2},$$

where $h_1 = h_1(u_1, u_2)$ and $h_2 = h_2(u_1, u_2)$ are functions of the coordinates u_1 and u_2 . We further express β and ξ as functions of the coordinates u_1 and u_2 ,

$$\beta = \beta(u_1, u_2),$$

$$\xi = \xi(u_1, u_2).$$

Often the surface current density $\bar{G}(u_1, u_2)$ in the aperture is expressed as a certain constant current I_0 divided by a certain constant length w and multiplied by a normalized, dimension-free surface current distribution function $\bar{g}(u_1, u_2)$,

$$\bar{G}(u_1, u_2) = \frac{I_0}{w} \bar{g}(u_1, u_2).$$

We also introduce a normalized, dimension-free electric field strength $\bar{e}(\theta, \phi)$ in the far zone field defined by the equation

$$\bar{E}(r, \theta, \phi) = \zeta I_0 \frac{e^{-ikr}}{r} \bar{e}(\theta, \phi),$$

where, as usual, $\zeta = \sqrt{\mu/\epsilon}$ is the intrinsic impedance of the free space. We then have

$$e_\theta = \frac{kb \sin \theta}{4\pi} \sum_{n=-\infty}^{\infty} \frac{j^{n+1} e^{-in\phi} S_n(ka \sin \theta, kb \sin \theta)}{H_n^{(2)}(ka \sin \theta)} \cdot \left[\frac{n \cos \theta}{kb \sin^2 \theta} i_n^{(1)} - i_n^{(2)} \right],$$

$$e_\phi = \frac{kb}{4\pi} \sum_{n=-\infty}^{\infty} \frac{j^n e^{-in\phi} V_n(ka \sin \theta, kb \sin \theta)}{H_n^{(2)'}(ka \sin \theta)} i_n^{(1)},$$

where

$$i_n^{(\epsilon)} = \frac{1}{bw} \iint g_\epsilon(u_1, u_2) e^{j[k\xi(u_1, u_2) \cos \theta + n\beta(u_1, u_2)]} \cdot h_1(u_1, u_2) h_2(u_1, u_2) du_1 du_2.$$

It appears from the preceding expressions that when the surface current density is purely axial, *i.e.*, when $g_1(\phi, z) = 0$, then the radiated field will have no component in the ϕ -direction, *i.e.*, $E_\phi = 0$. On the other hand, the formulas also show that when the surface current density is purely circumferential, *i.e.*, when

$g_2(\phi, z) = 0$, then the radiated field will have a θ component and a ϕ component that are both nonvanishing except in certain directions, *i.e.*, in this case the radiated fields will in general be elliptically polarized.

An obvious way of establishing a current distribution on a circular cylindrical surface coaxial with a circular conducting cylinder is to coat the conducting cylinder with a dielectric or permeable material and then to establish the current distribution on the outside surface of this coating. The expression for the field radiated from the current distribution on this structure has been obtained by an analysis similar to the one made above for the special case with no coating present. However, since the resulting expressions are so complicated that their practical utility may be questioned, they are omitted here.

CONCLUSION

It has been the object of this paper to obtain an analytical expression for the field radiated from slots in and from wire antennas near conducting, circular cylinders under the assumption that the field in the slots or the current in the wires are given. All the work done in this report on the field radiated from various types of slots in circular cylinders is based on the paper by Silver and Saunders [4] regarding the field radiated from an arbitrary slot in a circular cylinder; this paper has further served as a model for all the work done here on the field radiated from wire antennas near circular cylinders. The derivation of the formulas in this report has, therefore, in principle been a fairly straightforward matter. However, due to the complexity of some of the physical problems solved here, the derivations have in several cases become rather involved.

The expressions for the far zone field in all the cases treated in this paper are series expressions containing cylinder functions. In general, very little information regarding the radiated field can be obtained by inspection of the formulas. For obtaining the desired information, numerical computations must be carried out. For cylinders having a diameter of the same order of magnitude as the wavelength, most of the field expressions derived in this report are well suited for numerical computations. Due to the complexity of the formulas and to the large amount of parameter values for which it will in general be desirable to have numerical computations carried out, the use of an electronic computer is strongly indicated.

ACKNOWLEDGMENT

This paper is a survey of one part of a more comprehensive investigation which the author undertook for the P.E.C. Corporation, Boulder, Colo., on a contract with the Martin Company, Littleton, Colo. The author's thanks are due to the Martin Company for permission to pub-

lish this paper and to Werner Koppl, principal engineer and head of the antenna group at the Martin Company, for having suggested the problems and for encouragement during the work. The author thanks Villey Toft Pedersen for permission to include his material on the inclined slot in this paper, and Mrs. Tove Larsen for her assistance in preparing the final manuscript.

ADDENDUM

When presenting this paper at the Toronto Symposium, the author was informed by Dr. George Sinclair that part of the material given here on the inclined slot with a sine-shaped aperture field is contained in an unpublished report by Sinclair, "The Distant Field of an Arbitrary Slot Antenna in a Cylinder," Antenna Laboratory, The Ohio State University, Columbus, 1950. This report also contains some numerical results, which will be very valuable as a partial check of the extensive electronic computer computations we are undertaking.

BIBLIOGRAPHY

- [1] H. L. Knudsen, "Radiation from Slots in Circular Cylinders," P.E.C. Corporation, Boulder, Colo.; 1958.
- [2] J. R. Wait, "A Survey of the Recent Literature on Slot Radiators," Natl. Bur. Standards, Boulder Colo., Rept. No. 5051; 1957.
- [3] J. R. Wait, "Electromagnetic Radiation from Cylindrical Structures," Pergamon Press, Inc., New York, N. Y.; 1959.
- [4] S. Silver and W. K. Saunders, "The external field produced by a slot in an infinite circular cylinder," *J. Appl. Phys.*, vol. 21, pp. 745-749; August, 1950.
- [5] V. T. Pedersen, "Udstrålet felt fra skrå slidseantenne i cirkulær cylinder (Radiated Field from Inclined Slot Antenna in Circular Cylinder)," Master's thesis, Technical University of Denmark, Copenhagen, Denmark; 1958.
- [6] P. S. Carter, "Antenna arrays around cylinders," *Proc. IRE*, vol. 31, pp. 671-693; December, 1954.
- [7] H. L. Knudsen, "Bidrag til teorien for antennesystemer med hel eller delvis rotationssymmetri (Contributions to the Theory of Antenna Systems with Complete or in Part Rotational Symmetry)," I kommission hos Teknisk Forlag, Copenhagen, Denmark; 1953.
- [8] E. Nielsen, "Udstrålet felt fra ringkvasigruppe omkring ledende eller dielektrisk cirkulær cylinder (Field Radiated from Ring Quasi-Array Around Conducting or Dielectric Cylinder)," Master's thesis, Technical University of Denmark, Copenhagen, Denmark; 1953.
- [9] J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., Inc., New York, N. Y.; 1941.

Impedance Properties of Complementary Multiterminal Planar Structures*

GEORGES A. DESCHAMPS†

Summary—Booker has shown that Babinet's principle, properly extended to electromagnetic fields, leads to a simple relation between the impedances of two planar complementary structures. A relation, which generalizes this result, is found between the impedance matrices of two complementary n -terminal structures.

This relation is applied to the particular n -terminal structures having n -fold symmetry and to those that are also self-complementary. In the latter case the impedance matrix is real and entirely determined by the number of terminals. It is therefore independent of the exact shape of the elements composing the structure and of the frequency. By connecting in groups the terminals of such a structure various impedance levels, all frequency independent and real, may be achieved.

Structures having their terminal pairs in different locations in the plane are also considered. A self-complementary two-port structure is found to be equivalent, from the impedance point of view, to a length of lossy transmission line having a characteristic impedance of 60π ohms.

INTRODUCTION

BABINET'S principle, generalized to electromagnetism, gives a relation between the fields scattered or diffracted by two complementary plane structures. The structures to which the principle applies are made of infinitely thin perfectly conducting sheets of arbitrary shape. Two such structures are called complementary if one is obtained from the other by exchanging the open and the conducting portions of the plane.

After obtaining this generalized principle, Booker¹ showed that it implies a precise relation between the impedances of a pair of complementary two-terminal structures measured between closely spaced terminals in the plane of the structure. For example, a narrow slot in a conducting plane, excited across its center, and the complementary narrow strip, forming essentially a dipole with a gap at the center [see Fig. 1(a)], have their impedances Z_1 and Z_2 related by

$$Z_1 Z_2 = \left(\frac{1}{2}\zeta\right)^2 \quad (1)$$

where ζ is the intrinsic impedance of the surrounding medium (in free space, or in air, $\frac{1}{2}\zeta$ is practically 60π ohms). The same result holds for any pair of complementary two-terminal structures.

The first problem considered in this paper is the generalization of (1) to structures that have more than two terminals. Fig. 1(b) is an example of two complementary three-terminal structures. Since the impedance prop-

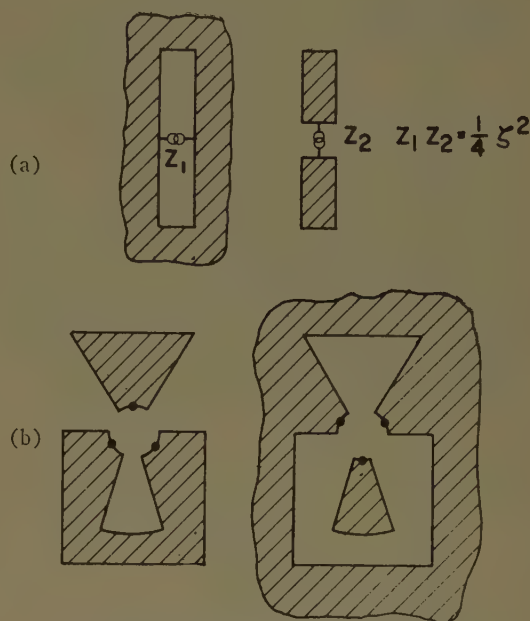


Fig. 1—Two-terminal and three-terminal complementary structures.

erties of these structures are now represented by matrices Z_1 and Z_2 it is readily seen that the relation between them must take a form different from (1).

Motivation for this investigation came from work done at the University of Illinois on frequency independent antennas. It had been noted by Mushiake and Rumsey² that Booker's relation implies that a two-terminal structure congruent to its complement, *i.e.*, "self-complementary," must have an impedance equal to 60π ohms independent of the frequency. Examples of such structures are shown in Fig. 2. For Fig. 2(a), made up of two right angles, the impedance can be computed³ directly and is indeed 60π ohms. The structure in Fig. 2(b) obtained by rotating the arbitrary curve C about O through angles multiple of 90° has also a frequency independent impedance of 60π ohms.

These considerations may seem of little practical importance since any self-complementary structure contains equal areas of conducting sheets and openings and must therefore be infinite. However, it was found by Rumsey,² and DuHamel and Isbell⁴ that for some structures of this type fed at the center the currents in the

* This work was supported by Contract AF33(616)-6079, Wright Air Development Center, Dayton, Ohio.

† University of Illinois, Urbana, Ill.

¹ H. G. Booker, "Slot aeriels and their relation to complementary wire aeriels," (Babinet's principle), *J. IEE*, pt. III-A, pp. 620-627; March-May, 1946.

² V. H. Rumsey, "Frequency independent antennas," 1957 IRE NATIONAL CONVENTION RECORD, pt. 1, pp. 114-118.

³ R. L. Carrel, "The characteristic impedance of two infinite cones of arbitrary cross section," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-6, pp. 197-201; April, 1958.

⁴ R. H. DuHamel and D. E. Isbell, "Broadband logarithmically periodic antenna structures," 1957 IRE NATIONAL CONVENTION RECORD, pt. 1, pp. 119-128.

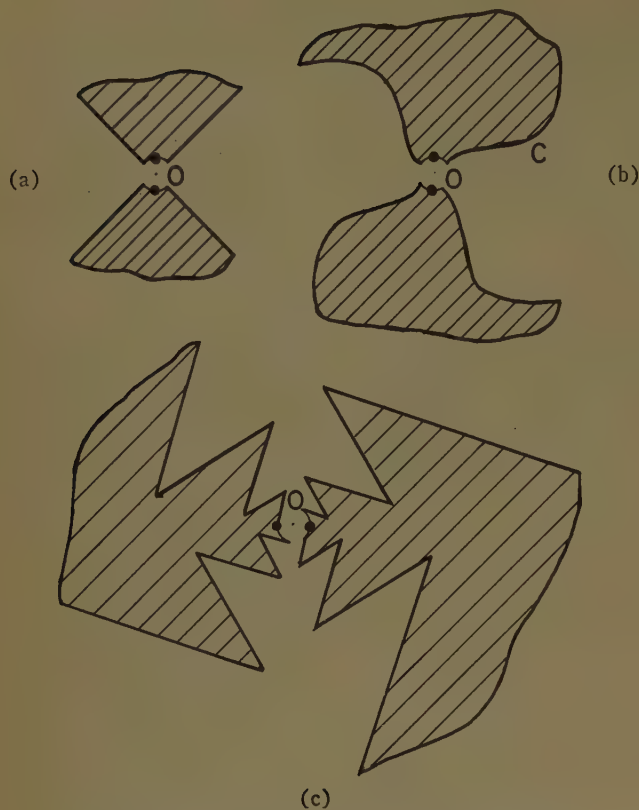


Fig. 2—Self-complementary two-terminal structures.

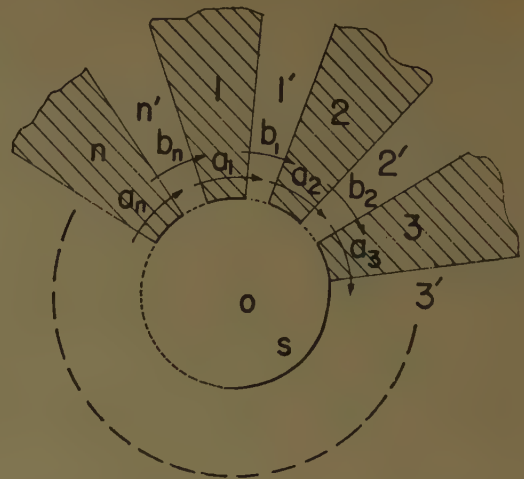
conducting sheets, and the electric fields tangent to the complementary openings, decrease rapidly with distance. The exact law of decrease has not been found but it is faster than $1/r$. As a consequence the structure may be truncated at a finite distance without affecting the input impedance. Fig. 2(c) is an example of such a structure. The impedance measured at O is found to be close to 60π ohms down to a frequency where the "end effect" becomes noticeable. (This occurs when the distance across the outside teeth approaches half a wavelength.)

Isbell and Mayes considered structures made up of several conducting sectors of this type. By connecting the terminals in groups they were able to obtain frequency independent impedances with values different from 60π .

One purpose of this paper is to show that the measured values can be predicted by using a proper extension of Booker's relation. The extension makes it possible to compute exactly the impedance matrix of a self-complementary n -terminal structure having n -fold symmetry. This result also gives, through a simple transformation, the characteristic impedance matrix for TEM wave propagation along symmetrical cylindrical and conical structures. Planar structures with terminal pairs at various locations in the plane are also briefly considered.

GENERALIZATION OF BOOKER'S RELATION

Consider the feed region of an n -terminal structure where n conducting sectors come together at a point O (see Fig. 3). Assume that these sectors are limited to the

Fig. 3—Feed region of the n -terminal structure.

outside of a sphere S small compared to the wavelength of operation. A source located inside S may be connected in several manners to these terminals producing different field configurations about the structure. Since the sphere is small we may describe the various possible connections as we would for a low-frequency circuit without regard for the exact shape of the conducting leads. The situation is completely specified by indicating which groups of terminals are connected to the two terminals of the source. For each such grouping a definite field configuration will result and a definite impedance will be seen by the source.

The method of solution will consist of associating those field configurations produced about complementary structures that are related by duality. From their comparison a relation will result between the corresponding voltages and currents at the terminals and therefore between the impedance matrices of the two structures.

Referring to Fig. 3, consider the various directed paths of integration designated by $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n$. They are drawn in the region I above the plane of the structure, very close to it, and have their beginning and end points in the plane. The b paths go from metal to metal and the a paths from opening to opening.

For any path c and any vector field U , let us introduce the shorthand notation $c \cdot U$ to indicate the integral of the vector U along the directed path c .

$$c \cdot U = \int_c U \cdot ds. \quad (2)$$

Let $F = (E, H)$ be an electromagnetic field produced about the given structure by some configuration of sources inside S .

The voltage difference between terminals i and $i+1$ is the integral of the vector E along the path b_i

$$V_i - V_{i+1} = b_i \cdot E. \quad (3)$$

The current I_i flowing into terminal i may be expressed by

$$I_i = 2a_i \cdot H. \quad (4)$$

This is seen by noting that the field F has even symmetry with respect to the plane of the structure and that a_i and its image by reflection in the plane (with orientation reversed) form a loop enclosing the i th terminal.

The integration paths a and b may be somewhat distorted in the region of the feed point without changing the values of the integrals (3) and (4).

Consider now the complementary structure obtained by replacing the open portions of the plane by conducting plates and replacing the metal by apertures. An acceptable solution for the electromagnetic field about this structure is obtained by taking the dual of F on one side of the plane (Region I for example) and the negative of the dual on the other side (Region II). This gives a field that has even symmetry with respect to the plane and satisfies the new boundary conditions.

The dual of a field $F=(E, H)$ is defined by

$$F' = (E', H') = (-\zeta H, \eta E) \quad (5)$$

where ζ and $\eta = \zeta^{-1}$ are respectively the intrinsic impedance and the intrinsic admittance of the surrounding space. It is a simple matter to verify that F' satisfies Maxwell's equation when F does.

For the field equal to F' in I and to $-F'$ in II relations similar to (3) and (4) will hold

$$I'_i = 2b_i \cdot H'. \quad (6)$$

$$V'_i - V'_{i+1} = a_{i+1} \cdot E'. \quad (7)$$

They define a set of currents and voltages that may exist at the terminals of the complementary structure and can be produced by a proper arrangement of sources in S .

Making use of (5) these may be expressed as

$$I'_i = 2\eta(V_i - V_{i+1}), \quad (8)$$

$$V'_i - V'_{i+1} = \frac{1}{2}\zeta I_{i+1}. \quad (9)$$

(By convention in these formulas as well as in (3), (4), (6), (7), $n+1$ is taken as equal to 1.) Formulas (8) and (9) may be collected in matrix form by introducing the vectors V, I, V', I' having for coordinates $(V_i), (I_i), (V'_i), (I'_i)$, respectively, and the matrix

$$\Delta = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & -1 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}. \quad (10)$$

Then (9) and (8) become:

$$\Delta V' = \frac{1}{2}\zeta I, \quad (11)$$

$$\Delta^T V = \frac{1}{2}\zeta I'. \quad (12)$$

(Δ^T denotes the transpose of Δ .)

These relations are general in the sense that to a condition specified by V and I on one structure is associated another condition described by V' and I' on the complementary structure.

In order to proceed we have to express the relations between V and I and those between V' and I' which describe the properties of the two structures. We shall only consider the case of truly n -terminal structures, i.e., those that may be fed arbitrary currents $I=(I_1, I_2, \dots, I_n)$ with the only restriction that $\sum I_k = 0$. The field configuration about the structure then depends upon $n-1$ independent parameters. When these are given, the voltage difference between any pair of terminals is defined and depends linearly on the vector I . Instead of choosing one of the terminals as a zero reference for the voltage it is convenient to use n voltage parameters $V=(V_1, V_2, \dots, V_n)$ related by the condition $\sum V_k = 0$. An impedance matrix may then be constructed such that

$$V = ZI \quad (13)$$

operates in the $n-1$ -dimensional space P defined by the relation

$$\sum_{k=1}^{k=n} x_k = 0 \quad (14)$$

where the x_k are the coordinates of a point.

A similar impedance matrix Z' describes the complementary structure

$$V' = Z'I'. \quad (15)$$

Starting from a given vector I in space P the voltage vector V results from (13). Then the vector $I' = 2\eta\Delta^T V$ represents a set of currents feeding the complementary structure and producing the voltage $V' = Z'I'$. But from relation (11), $I = 2\eta V'$. Finally

$$\Delta Z' \Delta^T Z I = \frac{1}{4}\zeta^2 I \quad (16)$$

for any vector in P .

This will be expressed by

$$\Delta Z' \Delta^T Z \equiv \frac{1}{4}\zeta^2. \quad (17)$$

The sign \equiv is to remind one that the two sides are equivalent only when applied to vectors in P . The left hand side transforms any vector into a vector belonging to P and therefore could not be equal to the right hand side without this restriction.

Eq. (17) is the generalization of Booker's relation to n -terminal structures. It will now be applied to symmetrical structures and then to self-complementary structures.

STRUCTURES WITH n -FOLD SYMMETRY

Let us now assume that the structure has n -fold rotational symmetry. This means that a rotation through the angle $\theta = 2\pi/n$ carries the structure upon itself. For example, Fig. 4 shows two complementary structures

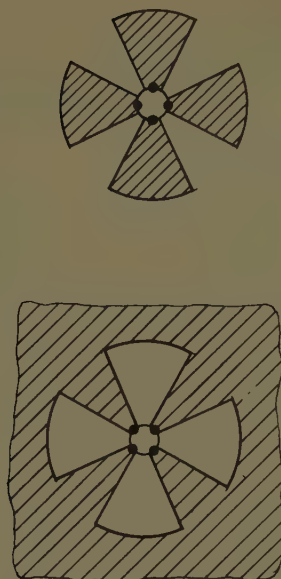


Fig. 4—Two complementary structures having four-fold symmetry.

with 4-fold symmetry. The corresponding matrix Z (and Z' for the complementary structure) will be completely determined by its first row. The next row is obtained by shifting all the elements one step to the right and taking the last element to the first place. The following rows are obtained by the same method and we may write:

$$Z = \begin{bmatrix} Z_0 & Z_1 & Z_2 & \cdots & Z_{n-1} \\ Z_{n-1} & Z_0 & Z_1 & \cdots & Z_{n-2} \\ Z_{n-2} & Z_{n-1} & Z_0 & \cdots & Z_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Z_1 & Z_2 & Z_3 & \cdots & Z_0 \end{bmatrix}. \quad (18)$$

Because of the symmetry of the matrix Z , $Z_{n-1}=Z_1$, $Z_{n-2}=Z_2$, \cdots and the number of parameters is actually $(n+1)/2$ for n odd and $n/2+1$ for n even.

Rather than using matrix notation it is convenient to consider Z as a sequence of n numbers

$$Z = (Z_0 Z_1 \cdots Z_{n-1}).$$

Similarly, V and I are sequences of n numbers

$$V = (V_0 V_1 V_2 \cdots V_{n-1}),$$

$$I = (I_0 I_1 I_2 \cdots I_{n-1}).$$

(By convention, the index n is equivalent to zero or, more generally, any index is defined modulo n .)

The relation between V and I becomes

$$V_i = \sum_k Z_{i-k} I_k \quad (19)$$

and is then expressed as a convolution

$$V = Z * I. \quad (20)$$

In order to represent the product by the matrix Δ as a convolution we introduce the sequence

$$U = (1 \ 0 \ 0 \ \cdots \ 0)$$

which plays the role of unity for the convolution product ($U * X = X$ for any sequence X) and the sequence

$$S = (0 \ 1 \ 0 \ \cdots \ 0).$$

Convolution of any sequence X by S has the effect of shifting each element of X by one step to the right and bringing the last element to the first place.

Multiplication by Δ then becomes convolution by $U - S$. Introducing also the sequence

$$\tilde{S} = (0 \ 0 \ \cdots \ 0 \ 1)$$

$\tilde{S}*$ operates a shift by one step to the left and multiplication by Δ^T becomes convolution by $U - \tilde{S}$.

The basic relation (17) between Z and Z' becomes

$$(U - S) * Z' * (U - \tilde{S}) * Z \equiv \frac{1}{4} \delta^2 \quad (21)$$

or commuting and reducing the factors, making use of the fact that $S * \tilde{S} = U$,

$$(2U - S - \tilde{S}) * Z' * Z \equiv \frac{1}{4} \delta^2. \quad (22)$$

[As for (17), this has to hold only when applied to a sequence I such that $\sum I_k = 0$.]

The usual technique for handling an equation of this type is to apply a Fourier transformation which will convert the convolution into an ordinary product. In the case of finite sequences this is also known as finding the "symmetrical components" of the sequence.

Introducing $\epsilon = \exp 2\pi j/n$ and the matrix

$$T = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \epsilon & \epsilon^2 & \cdots & \epsilon^{n-1} \\ 1 & \epsilon^2 & \epsilon^4 & \cdots & \epsilon^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \epsilon^{n-1} & \epsilon^{2(n-1)} & \cdots & \epsilon^{(n-1)^2} \end{bmatrix}. \quad (23)$$

The transform of a sequence

$$X = (X_0 X_1 \cdots X_{n-1})$$

is the sequence

$$x = (x_0 x_1 \cdots x_{n-1}) \quad (24)$$

obtained by

$$x = TX$$

where x and X are considered once more as column vectors rather than sequences. The inverse transformation is

$$X = T^{-1}x = \frac{1}{n} T^*x. \quad (25)$$

The asterisk means the complex conjugate. We shall systematically denote the transform of a sequence by the lower case letter corresponding to the capital letter describing the given sequence. Thus the transform of U is the sequence

$$u = (1, 1, \cdots 1)$$

and the transforms of S and \tilde{S} are, respectively,

$$s = (1, \epsilon, \dots, \epsilon^{n-1})$$

$$s^{-1} = (1, \epsilon^{-1}, \dots, \epsilon^{-(n-1)}).$$

Eq. (22) becomes

$$(2u - s - s^{-1})zz' = \frac{1}{4}\zeta^2. \quad (26)$$

Projecting this relation on the space P simply amounts to neglecting the zero component in the equality. Eq. (26) becomes

$$z_m z_m' \sin^2 \frac{m\pi}{n} = \left(\frac{1}{4}\zeta\right)^2 \quad (27)$$

for all $m \neq 0$.

This is the complementarity relation for symmetrical structures. It implies that if the impedance properties of a symmetrical structure are known, those of the complementary structure can be determined. Each symmetrical component or eigenvalue of the impedance matrix satisfies a relation similar to the original Booker's relation modified by a factor $\sin^2 m\pi/n$ depending on the order of the component and the number of terminals.

SELF-COMPLEMENTARY SYMMETRICAL STRUCTURE

A self-complementary symmetrical structure (with n -fold symmetry) may be obtained as shown in Fig. 5. Starting from a curve C_0 extending from the origin O to infinity, rotations of π/n about O bring it successively in positions $C_0', C_1, C_1', C_2, \dots, C_{n-1}, C_{n-1}'$ (C_n coincides with C_0). If the alternate sectors C_i to C_i' are filled with conducting plates, the structure obtained will have n -fold symmetry since a rotation of $2\pi/n$ brings it onto itself, and it will be self-complementary since a rotation of half that angle transforms it into the complementary structure.

By choosing for C_0 a curve with some oscillations in it or taking a zigzag line as was done by Isbell, a structure is obtained with small end effects. Impedance measurements may be taken on a truncated structure and they should agree with those taken on the infinite structure.

In the relation (27), $z_m' = z_m$, therefore

$$z_m = \frac{\frac{1}{4}\zeta}{\sin\left(\pi \frac{m}{n}\right)}, \quad m \neq 0. \quad (28)$$

The symmetrical components of the admittance sequence may be taken as

$$y_m = 4\eta \sin \pi \frac{m}{n}, \quad m \neq 0. \quad (29)$$

(As noted above we may assume $z_0 = y_0 = 0$ since both the zero order symmetrical components of V and I have been assumed equal to zero.)

By using the inverse transformation (25) the components of Y may be evaluated.

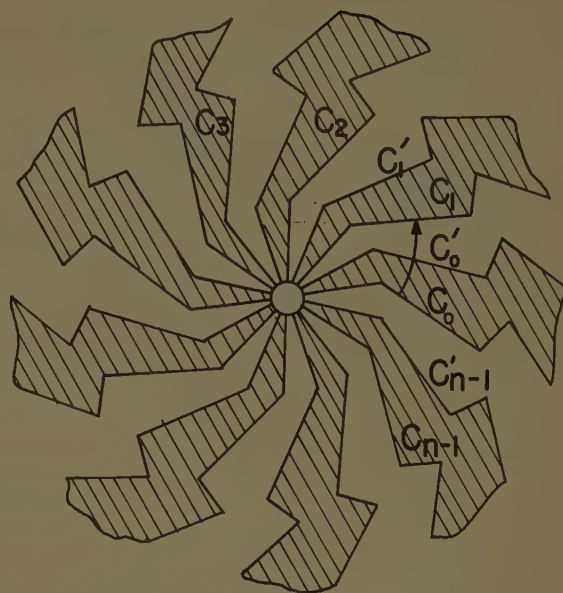


Fig. 5—Self-complementary symmetrical structure (nine-fold symmetry).

After some computation it is found that

$$Y_m = \frac{4\eta}{n} \frac{\sin \frac{\theta}{2}}{\cos m\theta - \cos \frac{\theta}{2}}, \quad (30)$$

where $\theta = 2\pi/n$ is the angle of one sector of the structure. Only the coefficient Y_0 is positive; all the others are negative and they add up to $-Y_0$ since $y_0 = 0$.

It should be noted that the z_m and y_m are also the eigenvalues of the matrices Z and Y belonging to the eigenvectors $I_{(m)}$ or $V_{(m)}$ represented by the m th column of the matrix T .

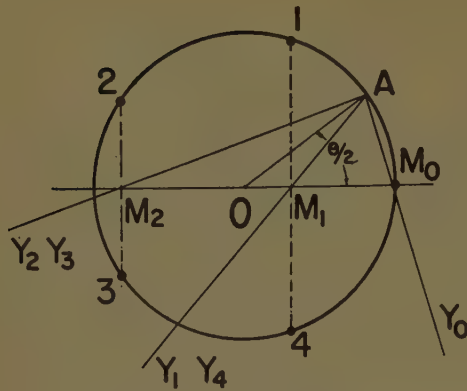
The formula (30) for Y_m has a simple graphical interpretation which may be useful to see how the coefficients of Y vary with n and m . If a circle is divided in n equal parts (see Fig. 6 where we have assumed $n=5$) the values of $\cos m\theta$ are read on the x axis as OM_m . Considering the point A at angle $\theta/2$ on the circle, the slope of the line AM_m is proportional to Y_m (more exactly the slope equals $-nY_m/4\eta$).

The admittance matrix, or the impedance matrix, of a symmetrical self-complementary structure is entirely defined by the number n of terminals. The coefficients are real and independent of frequency.

NUMERICAL RESULTS—EXPERIMENTAL VERIFICATION

When the admittance sequence (Y_m) is known, the impedance properties of any combination of terminals can be computed by simple circuit analysis techniques.

A systematic procedure can be found to deduce first the admittance matrix resulting from a grouping of the n terminals into p sets of connected terminals. If $C = (C_{ij})$ is the $p \times n$ connection matrix defined by $C_{ij} = 1$

Fig. 6—Construction for the coefficients $Y_{m/q}$ (case $n=5$).

when terminal i belongs to set j and $C_{ij}=0$ otherwise, the reduced admittance matrix is

$$Y_G = C^T Y C. \quad (31)$$

If the source is connected between group j and group k , all the I 's are zero except I_j and $I_k = -I_j$. The voltages are unknown except for the difference $V_j - V_k$. The equations are in sufficient number to define the ratio of $V_j - V_k$ to I_j which is the impedance sought.

The computations have been carried out for a number of configurations involving up to 7 elements and the results compared to experimental measurements.

The measurements are difficult because the feed lines of finite dimensions always disturb the ideal geometry. The thickness of the metal plate is also an important factor. In view of this, the agreement with observed values may be considered as satisfactory.

A plot of measured impedances obtained by D. Isbell and W. Guffey vs computed values (see Fig. 7) leads to the following observations. The experimental values are systematically below the theoretical ones. This may be explained by the finite thickness of the plate and in fact the agreement becomes better for thinner sheets. The percent error for a given thickness increases almost linearly with the number of terminals, independently of the manner in which they are connected.

Disagreement with the theoretical, real, and frequency independent value of the impedance is accompanied by a small variation of the impedance with frequency about a point on the real axis. This variation is of the same order of magnitude as the disagreement. For log-periodic structures the variation is periodic over an approximately circular locus. The values used in Fig. 7 are average impedances corresponding to the center of these circles.

Theoretical values for some of the configurations considered are represented graphically in Fig. 8. It is clear that by increasing the number of terminals, a large range of frequency independent values can, in principle, be obtained.

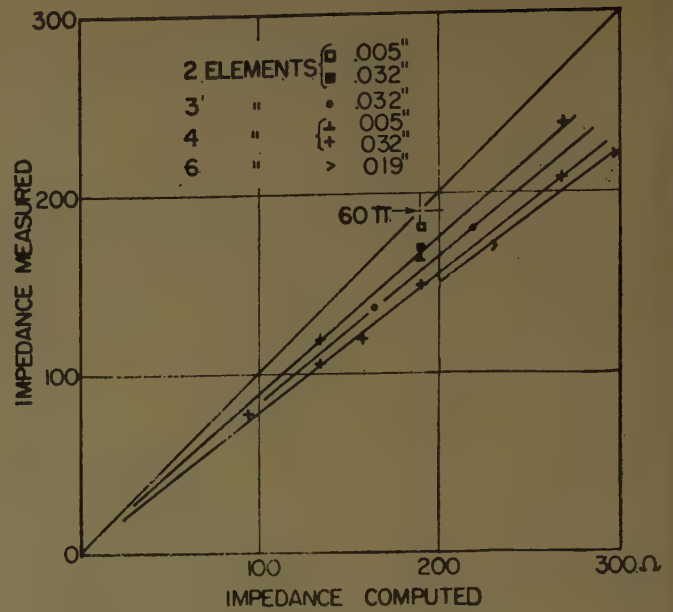


Fig. 7—Comparison with experiment.

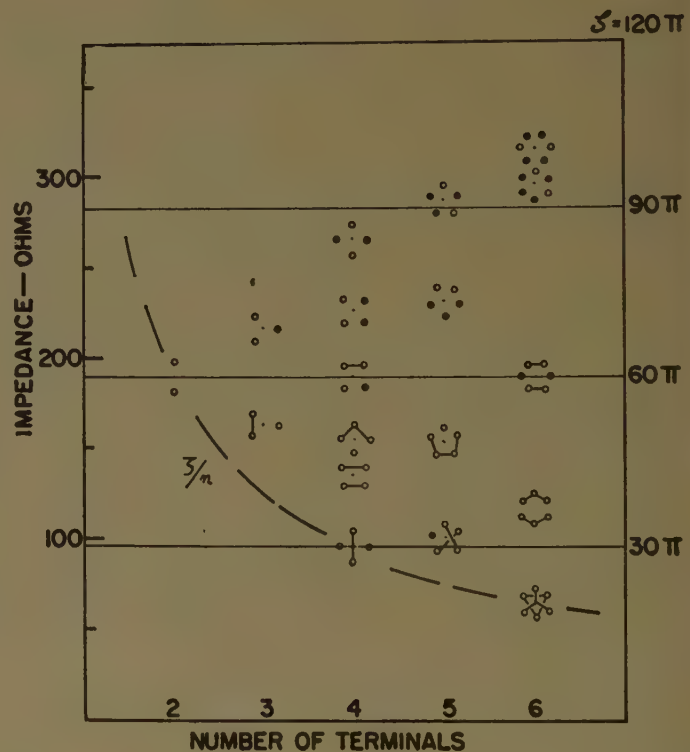


Fig. 8—Table of impedance levels obtained for various configurations of terminals. For each configuration, the two groups of terminals connected to the source are represented by small circles. The floating terminals are represented by black dots.

APPLICATION TO SOME ELECTROSTATIC AND TEM PROPAGATION PROBLEMS

Eqs. (17), (22), and (24)–(29) have been derived without reference to the particular shape of the elements composing the structure. They do of course apply when these elements are simple angular sectors limited by straight lines [see Fig. 9(b)]. It is known, however,

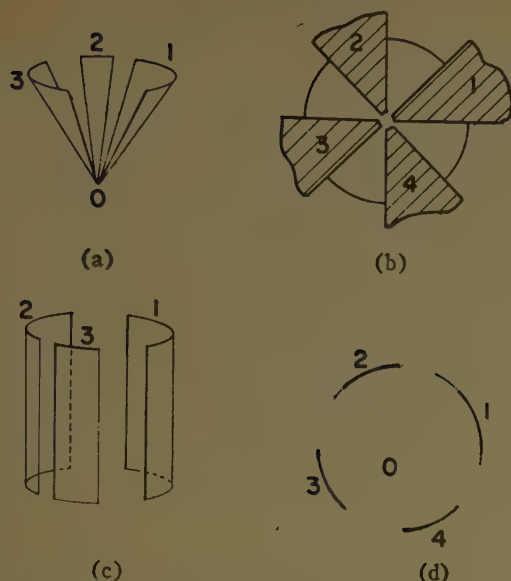


Fig. 9—Related electrostatic and TEM propagation problems.

that these special conical structures support TEM modes of propagation and the admittance matrix of the infinite structure then becomes the characteristic admittance matrix.

Furthermore the characteristic admittance is simply related to the capacitance matrix of the trace of the structure on a sphere of center O .

$$\frac{Y}{\eta} = \frac{C}{\epsilon} \quad (32)$$

The complementary relations (17), (22), and (27)–(29) therefore have their counterparts for the capacitance of structures made of conducting arcs on a circle. For example, a structure made of equal conducting arcs of circles [see Fig. 9(d)] separated by equal openings, has a capacitance matrix representable by a sequence:

$$C_m = \frac{4\epsilon}{n} \frac{\sin \frac{\theta}{2}}{\cos m\theta - \cos \frac{\theta}{2}} \quad (33)$$

(Note that this assumes $\sum Q_m = 0$, hence does not give information about the capacitance of the whole structure connected at a given potential with respect to infinity.)

The circular structure may be placed on an arbitrary sphere and considered as the trace of a conical set of plates [Figure 9(a)]. The characteristic admittance matrix of this conical structure is the same as that of the planar structure from which it comes.³

Finally the circular structure may be thought of as the trace of a cylindrical set of plates [Fig. 9(c)] and the characteristic admittance matrix is again the same as for the planar structure.

The appropriate complementary formulas could have been proved directly for each of these structures but it is worthwhile to note the relations between these problems.

STRUCTURES WITH SEVERAL TERMINAL REGIONS

The structures considered so far had all their terminals coming to a point or in practice connected to sources inside a region small in terms of wavelength.

One may also consider structures having terminals at different locations in the plane. Fig. 10 shows an example of a five-terminal structure having two terminal regions. The terminals may be numbered (1, 2, 3) (4, 5). Those of the complementary structure will be (1', 2', 3') (4', 5') as shown in the figure. By convention $i+1$ is the terminal "next to" i , thus $3+1=1$, $5+1=4$. It is convenient to use as voltage parameter V_i , the potential difference between terminal i and $i+1$. The sum of the V_i as well as the sum of I_i is thus zero for every terminal region.

Introducing the shift operator defined by

$$S \begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \end{bmatrix} = \begin{bmatrix} I_2 \\ I_3 \\ I_1 \\ I_5 \\ I_4 \end{bmatrix}, \quad (34)$$

the relation between the impedance matrices of the two complementary structures becomes

$$Z'Z \equiv \frac{1}{4}\zeta^2 S. \quad (35)$$

There again the sign \equiv means that the two sides of the equation give the same result when applied to a vector I such that $\sum I_k = 0$. Eq. (35) is obtained by the same method as (17). This is an alternative form of complementarity which could have been used instead of (17). The only difference is in the choice of the voltage parameters. Those used in (17) were found more convenient in solving the problem of grouping of the terminals.

TWO-PORT SELF-COMPLEMENTARY STRUCTURES

A case of special interest is that of the two-port structure, having two terminal regions with two terminals each.

Choosing at each location, 1 and 2, a $+$ terminal, the currents

$$I = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \quad (36)$$

are defined as those flowing into 1₊ and 2₊, the voltages

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \quad (37)$$

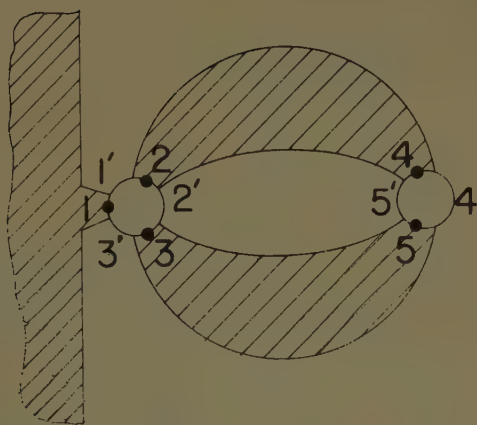


Fig. 10—Planar structure with 3+2 terminals.

are respectively the voltage differences between 1_+ , 1_- and 2_+ , 2_- . The impedance matrix Z relates V and I .

$$V = ZI. \quad (38)$$

Fig. 11 shows a self-complementary two-port structure. By reflection into the straight line this structure is transformed into its complement.

Applying the duality transformation to the field as was done in (6)–(9) it is seen that

$$\begin{aligned} V_1' &= \frac{1}{2}\xi I_1, \\ I_1' &= 2\eta V_1, \end{aligned} \quad (39)$$

while

$$\begin{aligned} V_2' &= -\frac{1}{2}\xi I_2, \\ I_2' &= -2\eta V_2. \end{aligned} \quad (40)$$

The sign reversal comes from the fact that in the duality transformation each quantity (V' or I') is related to the dual quantity (I or V) belonging to the terminal immediately to its left (seen from the region above the plane). At location 1 in Fig. 11, this relates the two + terminals, but at location 2 it relates the + terminal to the - terminal of the reflected structure.

Introducing a matrix

$$\sigma = \begin{pmatrix} + & 0 \\ 0 & - \end{pmatrix} \quad (41)$$

the formulas (39) and (40) may be expressed as

$$\begin{aligned} V' &= \frac{1}{2}\xi\sigma I \\ I' &= 2\eta\sigma V, \end{aligned} \quad (42)$$

and using (38) which applies also to V' , I' ,

$$\frac{1}{2}\xi\sigma I = Z2\eta\sigma V \quad (43)$$

or, finally,

$$(\sigma Z)^2 = \frac{1}{4}\xi^2. \quad (44)$$

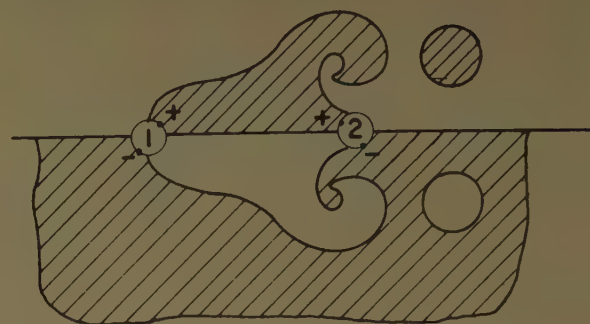


Fig. 11—Self-complementary two-port structure.

This is the relation that Z must satisfy in order to represent the self-complementary structure.

Explicitly, if

$$Z = \begin{vmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{vmatrix}, \quad (45)$$

it follows that

$$\begin{cases} Z_{11} = Z_{22} \\ Z_{11}^2 - Z_{12}^2 = \frac{1}{4}\xi^2. \end{cases} \quad (46)$$

This may be recognized as the impedance matrix of an ideal attenuator having a characteristic impedance of 60π ohms. The attenuation and phase shift through the element cannot be found from the symmetry considerations but depend on the form of the structure.

Another method of proving the equivalence with an attenuator is to consider the transformation of impedance (or reflection coefficient) through the two-port. If a resistive load of 60π ohms terminates 2, an impedance of 60π ohms will be seen at 1. Plotting 60π at the center O of the reflection chart as in Fig. 11, the point O becomes its own image (iconocenter of the transformation). If an open-circuit load P is mapped at point P' , the short circuit load Q will be mapped at Q' corresponding to the reciprocal impedance with respect to 60π . The segment $Q'P'$ has therefore O as its middle point. The image of the unit circle Γ is a concentric circle Γ' . An equivalent circuit for the structure is therefore a length of transmission line with 60π ohm characteristic impedance in cascade with an ideal attenuator.

The Bandwidth of Helical Antennas*

T. S. M. MACLEAN† AND R. G. KOUYOUMJIAN‡

Summary—The propagation along a finite helix is examined by using a propagation constant which was determined by Sensiper for the infinite helix. By utilizing the Hansen and Woodyard condition, the bandwidth variation with the dimensions of the helix is obtained. It is shown that for a pitch angle ψ equal to 13° , the normally accepted bandwidth of $0.77 < C_\lambda < 1.25$, where C_λ is the circumference in terms of free-space wavelengths, is applicable only to antennas up to about two wavelengths long. A curve showing the bandwidth as a function of length has been computed, and the shape of this curve has been confirmed experimentally. Good patterns and axial ratio with sidelobe levels no greater than -10 db were obtained for helices as long as 10 wavelengths. Good patterns and axial ratio may also be obtained for ψ as small as 2° provided that the helix is sufficiently long to establish the surface wave.

INTRODUCTION

THE discovery of axial radiation from a uniform helix by Kraus¹ in 1947 resulted in a thorough and fairly complete investigation of the helical antenna, from the viewpoint of both pattern and impedance, over a wide frequency range. This investigation covered antennas of between three and ten turns and pitch angles of 5° to 24° . At the same time, sufficient analysis was carried out to give an adequate understanding of the operation of the antenna in terms of inward- and outward-traveling waves along the conductor. These investigations were completed by 1950.

An alternative and more complete theoretical approach to the problem of electromagnetic wave propagation along helical conductors was given by Senisiper² in 1951, but since he was interested primarily in its application to traveling-wave tubes, and his approach was necessarily mathematical, little effort has been made to apply his results to the helical antenna. The object of the present investigation was to carry this out, and as a result, new information is available about the limits of the helical pitch angle which may be employed, and about the bandwidth of helical antennas as long as ten wavelengths.

PHASE VELOCITY

The characteristics of the helical antenna are functions of 1) the over-all length in free-space wavelengths (L_λ); 2) the pitch angle (ψ); 3) the circumference of the imaginary cylinder on which the helix is wound (C_λ); and 4) to a lesser extent, of the conductor diameter (d_λ). It is desirable to consider a single parameter, the

phase velocity v along the conductor, rather than each of these separately, since v is a function of each of these variables and is directly related to the radiation characteristics of the helical antenna. The same simplification can also be used profitably in the case of the Yagi antenna.³

The determinantal equation to be solved for obtaining numerical values of phase velocity along an infinite helix is²

$$0 \simeq \sum_m \left\{ \left(\beta_0^2 a^2 - k^2 a^2 + k^2 a^2 \frac{m^2 \cot^2 \psi}{(\gamma_m a)^2} \right) I_m(\gamma_m a) K_m(\gamma_m a) + k^2 a^2 \cot^2 \psi I_m'(\gamma_m a) K_m'(\gamma_m a) \right\} \frac{\sin m\pi \frac{\delta}{h}}{m\pi \frac{\delta}{h}} \quad (1)$$

where

- β_0 is the propagation constant to be determined,
- a = radius of imaginary cylinder on which the helix is wound,
- $k = \omega \sqrt{\mu \epsilon}$ the free-space phase constant,
- ψ = helical pitch angle = $\tan^{-1} h/2\pi a$,
- $\gamma_m^2 = (\beta_m^2 - k^2)$,
- $\beta_m = \beta_0 + m(2\pi/h)$,
- δ = tape width of helix conductor.

The primes on the modified Bessel functions denote differentiation with respect to the argument.

An alternative but essentially similar form of the equation has been given by Chodorow and Chu,⁴ and it is this form which has been used in the computations throughout:

$$0 \simeq U_0 + V_0 + L_0 \sin \psi \frac{1}{\left(\frac{\delta}{h} 2\pi \right)^2} \left[C_3(0) - C_3\left(\frac{\delta}{h} 2\pi \right) \right] + \sum_1 \left[\frac{\sin \left(m \frac{\delta}{h} \pi \right)}{\left(m \frac{\delta}{h} \pi \right)} \right]^2 \cdot \left(Y_m + Y_{-m} - L_0 \frac{\sin \psi}{m} \right) \quad (2)$$

* The research reported in this paper was sponsored in part by the U. S. Army Signal Engrg. Lab., Fort Monmouth, N. J.

† Antenna Lab., The Ohio State University, Columbus, Ohio. On leave of absence from Edinburgh University, Edinburgh, Scotland.

‡ Antenna Lab., The Ohio State University, Columbus, Ohio. ¹ J. D. Kraus, "Helical beam antenna," *Electronics*, vol. 20, pp. 109-111; April, 1947.

² S. Sensiper, "Electromagnetic Wave Propagation on Helical Conductors," Res. Lab. of Electronics, Mass. Inst. Tech., Cambridge, Tech. Rep. No. 194; May, 1951.

³ H. W. Ehrenspeck and H. Poehler, "A New Method for Obtaining Maximum Gain from Yagi Antennas," presented at the WESCON Convention, Los Angeles, Calif.; August, 1956.

⁴ M. Chodorow and E. L. Chu, "Cross Wound Helices for Traveling-Wave Tubes," Microwave Lab., Stanford University, Stanford, Calif., Rept. No. 249; November, 1954. Prepared under Office of Naval Res. Contract N6 onr 25123 (NR373 361).

where

$$\begin{aligned}U_0 &= (\gamma_0 a)^2 K_0(\gamma_0 a) I_0(\gamma_0 a) \sin^2 \psi, \\V_0 &= -(ka)^2 K_1(\gamma_0 a) I_1(\gamma_0 a) \cos^2 \psi, \\L_0 &= (\gamma_0 a)^2 \sin^2 \psi - (ka)^2 \cos^2 \psi,\end{aligned}$$

with

$$\begin{aligned}\gamma_m^2 &= \beta_m^2 - k^2, \\ \beta_m &= \beta_0 + m \frac{2\pi}{h},\end{aligned}$$

$$\begin{aligned}C_3\left(\frac{\delta}{h} 2\pi\right) &= \sum_{m=1}^{\infty} \frac{\cos m\left(\frac{\delta}{h} 2\pi\right)}{m^3} \\ &= 1.202 + \frac{1}{2}\left(\frac{\delta}{h} 2\pi\right)^2 \log\left(\frac{\delta}{h} 2\pi\right) \\ &\quad - \frac{3}{4}\left(\frac{\delta}{h} 2\pi\right)^2 - \frac{1}{288}\left(\frac{\delta}{h} 2\pi\right)^4 - \dots\end{aligned}$$

$$\begin{aligned}Y_m &= (\gamma_m a)^2 K_m(\gamma_m a) I_m(\gamma_m a) \sin^2 \psi \\ &\quad + \left[\left(\frac{m\beta_m a}{\gamma_m a} \right)^2 K_m(\gamma_m a) I_m(\gamma_m a) \right. \\ &\quad \left. + (ka)^2 K_m'(\gamma_m a) I_m'(\gamma_m a) \right] \cos^2 \psi \\ &\quad - (m\beta_m a) K_m(\gamma_m a) I_m(\gamma_m a) \sin^2 \psi.\end{aligned}$$

It is to be noted that in common with other periodic structures,⁵ there are certain forbidden regions associated with the helix, in which a solution to the above equations may not be found. The regions where solutions may be found correspond to

$$\gamma_m a = [(\beta_m^2 - k^2)a^2]^{1/2} > 0 \quad (3)$$

or

$$|\beta_m a| > ka.$$

For both positive and negative β_0 and $|m| \geq 1$, this condition becomes²

$$|m| + \frac{ka}{\cot \psi} < \frac{|\beta_0| a}{\cot \psi} < |m| - \frac{ka}{\cot \psi}. \quad (4)$$

A graph of this inequality is shown in Fig. 1, where the shaded areas indicate the forbidden regions for different values of m . The pass band region for the beam-mode type of helical antenna is also shown. It will be noted that the forbidden regions become continuous for $ka/\cot \psi > 0.5$, i.e., for $d > \lambda/2$. This condition also arises in the theory of linear end-fire arrays, where for $d > \lambda/2$ there is no longer a single major lobe in the pattern.

The choice of coordinates shown in Fig. 1 has the advantage of making the graph applicable to any value of

pitch angle ψ , which would not be true for the ω - β coordinates more commonly used with periodic structures.⁵ At the same time, the slope of the solution to the determinantal equation for a given value of ka , i.e.,

$$\frac{ka}{\cot \psi} / \frac{\beta_0 a}{\cot \psi},$$

still gives the axial phase velocity along the helix as it does when the ω - β coordinates are used.

THE BANDWIDTH OF A MEDIUM-PITCH HELIX

The reason for obtaining numerical values of the propagation constant along an infinite helix was to determine if such values would be of use in predicting an upper frequency limit for a finite helical antenna. It is known, for example, that if the propagation constant along the conductor satisfies the Hansen-Woodyard condition along the axis, the pattern will be an optimum one for a uniform helix. If the phase velocity exceeds that for the Hansen-Woodyard condition, the width of the main lobe will be increased, while if it is lower, the sidelobe level will be increased. In order to examine the bandwidth of the medium-pitch helix quantitatively, the case of $\psi = 13^\circ$ will be considered.

The reasons for choosing $\psi = 13^\circ$ were that a) accurate experimental values of phase velocity for this pitch angle and for a specific length of 1.6λ were available, and b) it was known that an angle of about this size gives the largest bandwidth for a helix 1.6 wavelengths long.

Fig. 2 illustrates this graphically for an array of 50 isotropic elements. The array factor is

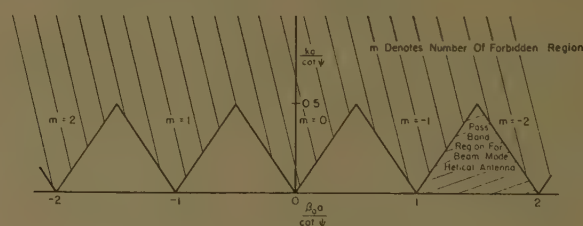


Fig. 1—Chart of forbidden regions.

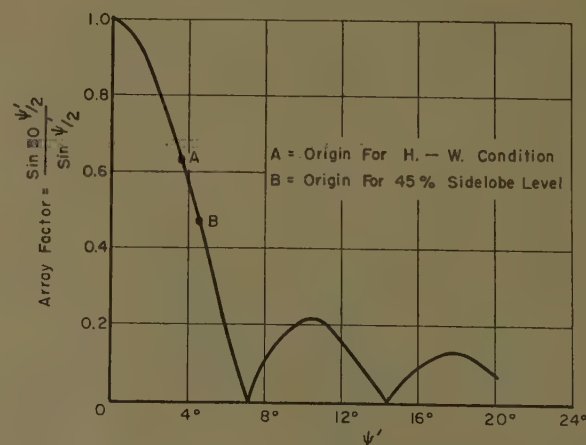


Fig. 2—Array factor for 50 isotropic elements.

⁵ D. A. Watkins, "Topics in Electromagnetic Theory," John Wiley and Sons, Inc., New York, N. Y.; 1958.

$$\frac{\sin 50\psi'/2}{\sin \psi'/2},$$

where

$$\begin{aligned}\psi' &= \beta d \cos \phi - \alpha \\ &= 2\pi C_\lambda \tan \psi \cdot \cos \phi - \frac{L_\lambda}{v/c} \cdot 2\pi.\end{aligned}\quad (5)$$

Only the part of this pattern to the right of the point corresponding to $\phi=0^\circ$ is traced out; this point is a function of the phase velocity v along the conductor. It should be noted that λ is the free-space wavelength. Point A represents the point $\phi=0^\circ$ for the Hansen-Woodyard phase condition, and point B is the point $\phi=0^\circ$ where the sidelobe level has increased to 45 per cent of the main beam. This sidelobe level will be arbitrarily taken to represent the upper allowable frequency of operation for the antenna.

It will be noted that in moving from A to B , ψ' has increased from $3.6^\circ (= \pi/50)$ to $4.6^\circ (= 0.080$ radian). Substituting these values in (5) and taking $C_\lambda = 1.0$ enable the two corresponding values of v to be calculated. The percentage change in v as A moves along the array factor curve to B is only 0.22 per cent. Use will be made of this result later.

Conversely, if v as a function of C_λ is known accurately, the value of C_λ for which the sidelobe level will be equal to any percentage of the main beam can be calculated from the same equation.

In solving the determinantal equation, the normal range of $ka (= C_\lambda)$ from 0.7 to 1.3 has been used. Values of the phase velocity over this range are shown in Fig. 3, where it must be remembered that it is the phase velocity along the conductor which is plotted. This is related to the fundamental axial phase velocity by

$$v_{\text{conductor}} = \frac{v_{\text{axial}}}{\sin \psi}.$$

Superimposed on the same graph is the Hansen-Woodyard condition for an infinitely long array, which also represents the in-phase field condition for a finite array. It is noteworthy that over the region $0.7 < ka < 1.0$ the two curves coincide within the thickness of the curve, although (a) is always slightly below (b). This result encourages one to utilize the solution for the infinite helix in connection with the helical-antenna problem.

However, in order to estimate the upper frequency limit of the finite helix, the Sensiper solution must be compared with the Hansen-Woodyard curve for the finite helix. Fig. 4 compares the Sensiper solution with the Hansen-Woodyard condition for 50 turns. Using the point where the divergence of the two curves becomes significant, an estimate for the upper frequency limit is that $C_\lambda \approx 1.1$. This can be evaluated more accurately using (5); i.e.,

$$2\pi \left(C_\lambda \tan \psi - \frac{L_\lambda}{v/c} \right) = -(2\pi + 0.080).$$

It is noted that $L_\lambda = C_\lambda \sec \psi$, so that for $\psi = 13^\circ$ there are two unknowns, C_λ and v . Using the relationship between v and C_λ resulting from Sensiper's solution, the desired solution of (5) is obtained. For the case considered here the solution is $C_\lambda = 1.10$, confirming the previous estimate of the same value.

This result has been obtained by use of only the array factor for 50 turns, and the calculated phase velocity for the infinite helix. In the same way, using the array factors for 3, 5, 10, 15, \dots , 45 turns, and the phase velocity for the infinite helix, the upper frequency limit C_λ can be computed for each of these numbers of turns. The results of these computations are shown in Fig. 5 as curve (a). For comparison, the experimental results for $3 \leq n \leq 50$ turns are shown as curve (b).

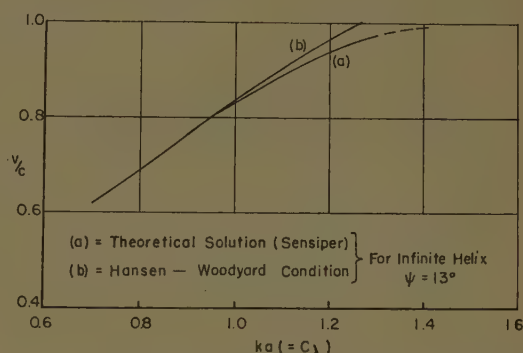


Fig. 3—Phase velocity vs circumference in wavelengths.

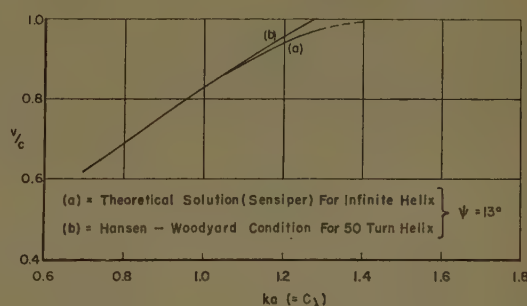


Fig. 4—Phase velocity vs circumference in wavelengths.

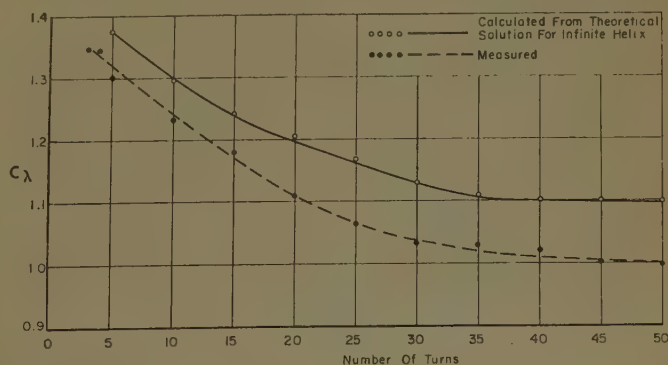


Fig. 5—Upper frequency limit vs number of turns for $\psi = 13^\circ$.

The agreement in the variation with number of turns is satisfying, although the calculated absolute values are necessarily high. This follows since the phase velocities used correspond to those for the infinite helix, which are known to be greater than the ones applying to a finite helix.

In order to complete the picture of the variation of bandwidth with the length of the antenna, it is necessary to consider also the variation, if any, of the lower frequency limit. The physics of the situation in this case is quite different from the limitation at the upper end. There the pattern break-up is caused by the phase velocity's not increasing rapidly enough with frequency, but at the lower end the pattern begins to become useful only with the effective launching of a new transmission mode. For a 13° pitch angle, this occurs at $0.77 C_\lambda$ for an infinite helix. Measured values of approximately $0.75 C_\lambda$ for helices as short as 0.7 wavelength suggest that this frequency remains substantially constant with length. Therefore, in Fig. 5 it is suggested that the ordinate of upper frequency limit essentially determines the ratio of upper to lower frequency limits with a scale conversion factor of $1/0.77$ or 1.3. This result corrects the earlier, quite widely held notion that the useful frequency range of a helical antenna centers in the frequency $C_\lambda=1.0$. According to the results reported here, however, it should rather be considered as existing above $C_\lambda=0.77$ for the 13° pitch angle case. The value of 0.77 is a fixed lower limit, while the upper limit is variable.

In obtaining the experimental results for Fig. 5(b), it was necessary to take patterns of E_θ , E_ϕ , and axial ratio for $3 \leq n \leq 50$ turns. A selection of these patterns is shown in Figs. 6-20. The most important of these are the ones for large n , since it is believed that these represent the patterns of the longest helices which have as yet been tested. In all cases it can be seen that both the patterns and the axial ratios are good. The measurement frequency of 8 kmc corresponds to $0.92 C_\lambda$.

The variation of half-power beamwidth with length of the antenna is shown in Fig. 21 for the E_θ and E_ϕ patterns. These curves are based on the experimental measurements shown in the previous figures, and agree closely with the values which may be predicted from Hansen-Woodyard conditions. The constant frequency to which the curves are applicable is $0.92 C_\lambda$.

At still higher frequencies corresponding to the upper bandwidth limit, the larger sidelobes are accompanied by a smaller half-power beamwidth which is plotted in Fig. 22. This occurs as the result of a reduction in phase velocity so that the array factor is shifted further into the imaginary region.⁶

In Figs. 23 and 24 the measured variation of axial ratio with length of the antenna is shown for the frequencies used in Figs. 21 and 22.

⁶ F. K. Goward, "An improvement in end-fire arrays," *J. IEE*, vol. 94, pt. III, pp. 415-418; November, 1947.

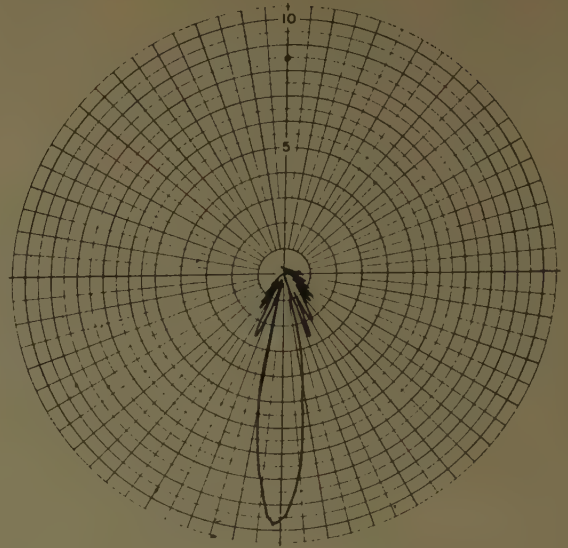


Fig. 6— E_θ pattern for 50-turn, 13° helix; $f=8$ kmc.

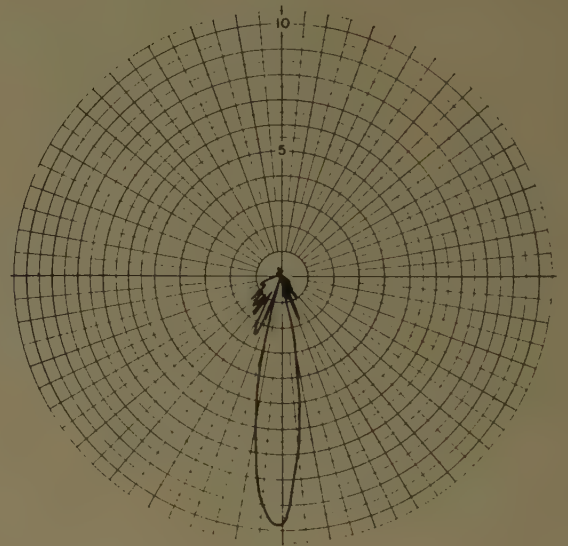


Fig. 7— E_ϕ pattern for 50-turn, 13° helix; $f=8$ kmc.

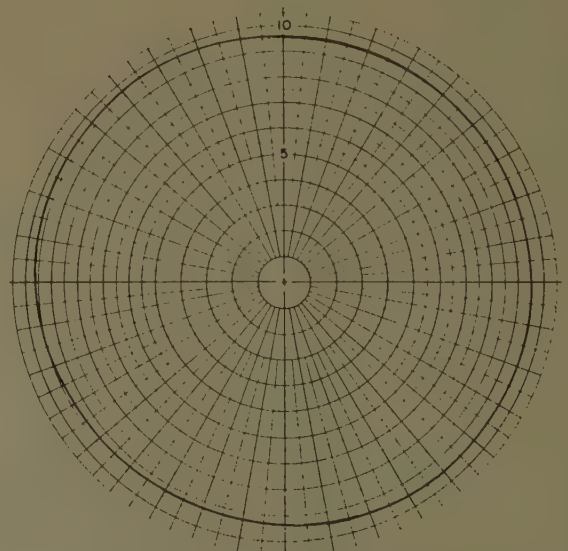


Fig. 8—Polarization pattern for 50-turn, 13° helix; $f=8$ kmc.

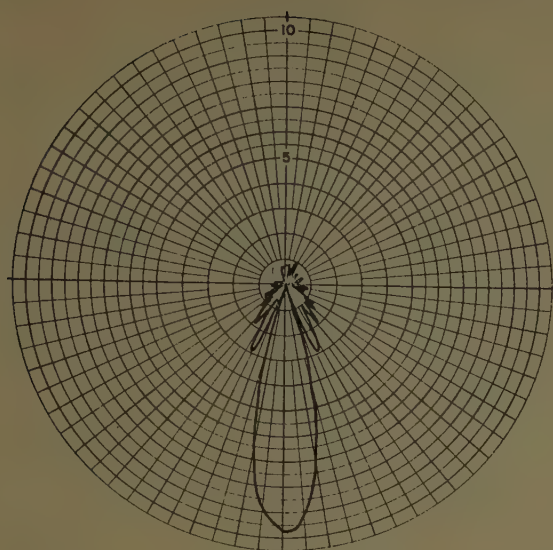


Fig. 9— E_θ pattern for 40-turn, 13° helix; $f=8$ kmc.

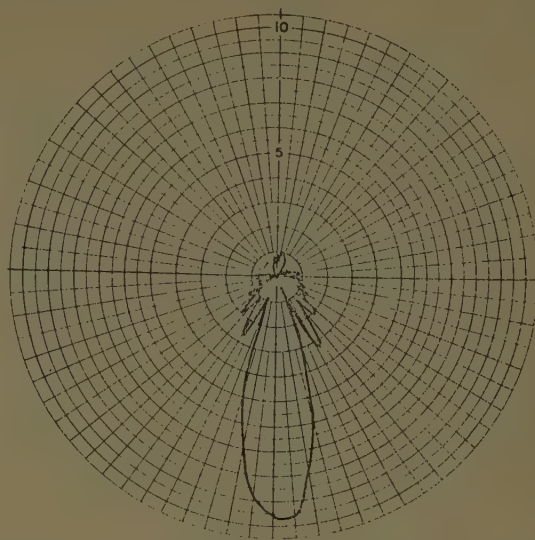


Fig. 12— E_θ pattern for 30-turn, 13° helix; $f=8$ kmc.

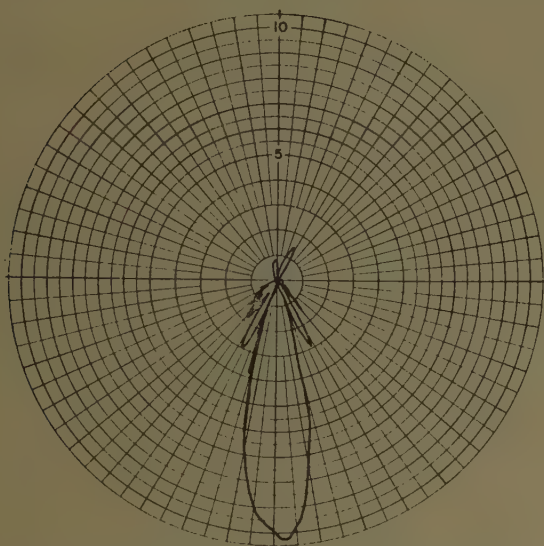


Fig. 10— E_ϕ pattern for 40-turn, 13° helix; $f=8$ kmc.

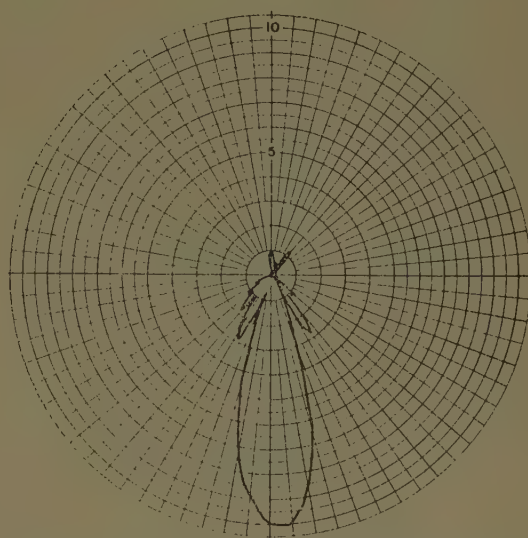


Fig. 13— E_ϕ pattern for 30-turn, 13° helix; $f=8$ kmc.

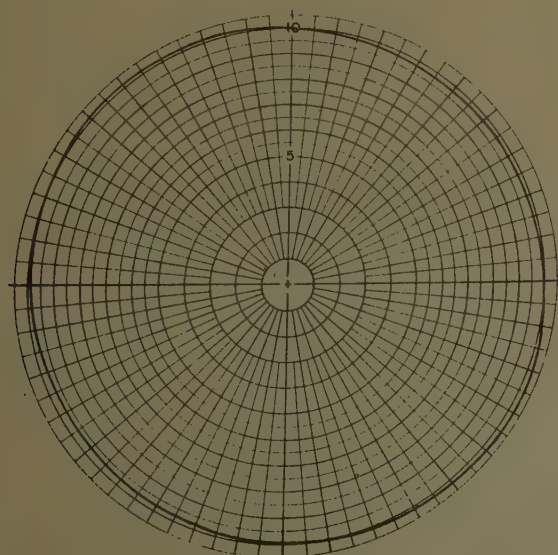


Fig. 11—Polarization pattern for 40-turn, 13° helix; $f=8$ kmc.

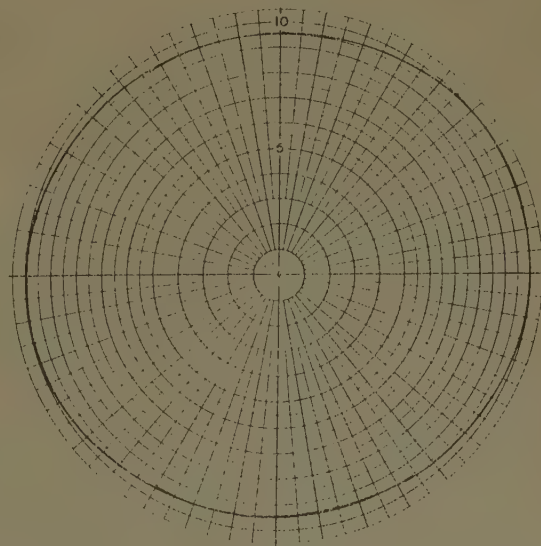
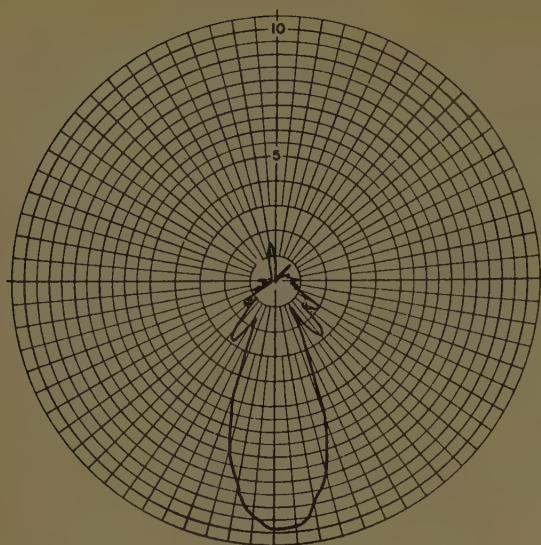
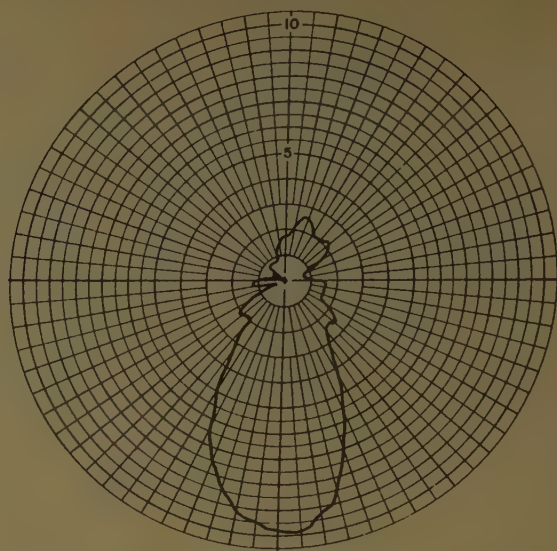
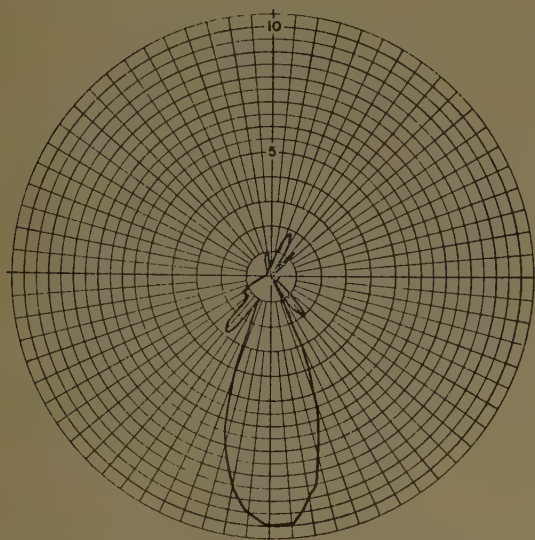
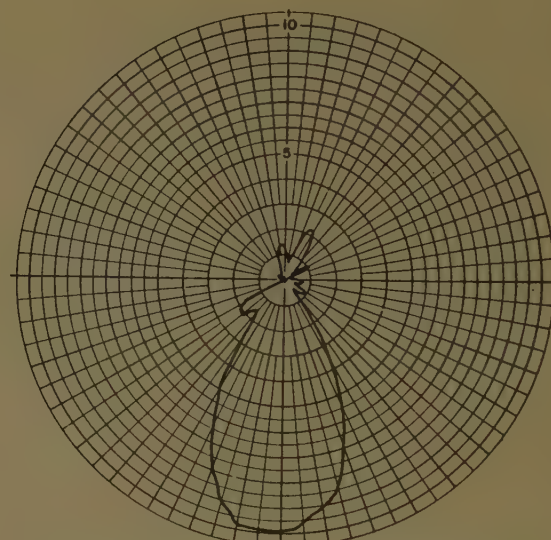
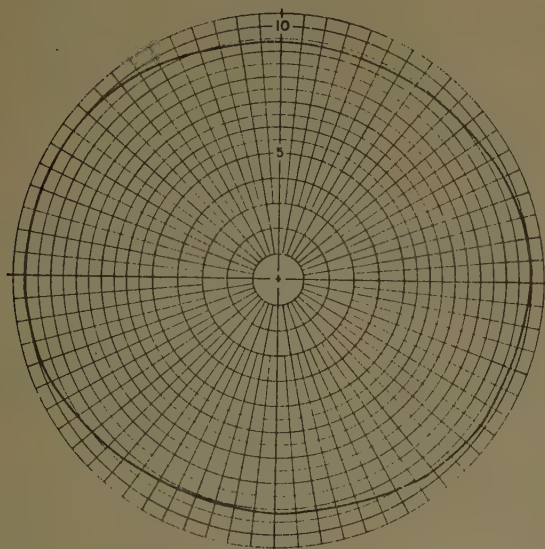
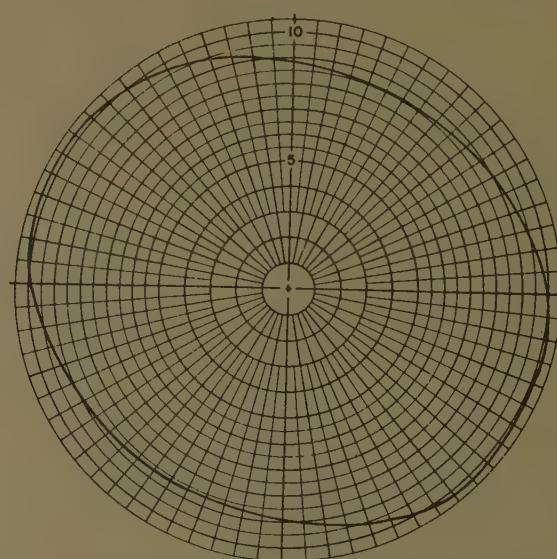


Fig. 14—Polarization pattern for 30-turn, 13° helix; $f=8$ kmc.

Fig. 15— E_θ pattern for 20-turn, 13° helix; $f=8$ kmc.Fig. 18— E_θ pattern for 10-turn, 13° helix; $f=8$ kmc.Fig. 16— E_ϕ pattern for 20-turn, 13° helix; $f=8$ kmc.Fig. 19— E_ϕ pattern for 10-turn, 13° helix; $f=8$ kmc.Fig. 17—Polarization pattern for 20-turn, 13° helix; $f=8$ kmc.Fig. 20—Polarization pattern for 10-turn, 13° helix; $f=8$ kmc.

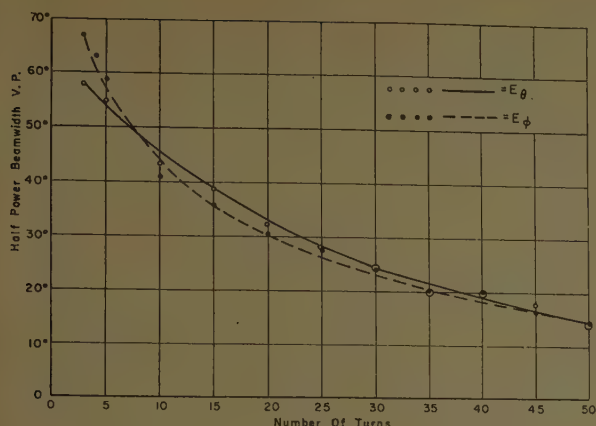


Fig. 21—Half-power beamwidth vs number of turns, $f=8$ kmc, 13° helix.

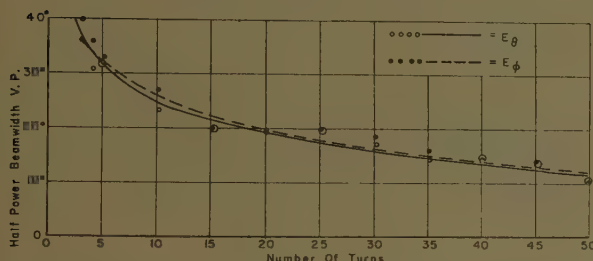


Fig. 22—Minimum half-power beamwidth vs number of turns, 13° helix.

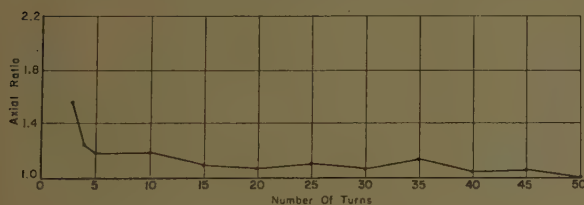


Fig. 23—Voltage axial ratio vs number of turns, 13° helix; $f=8$ kmc.

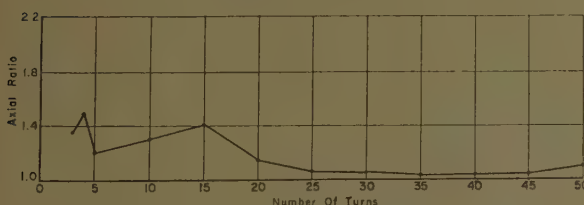


Fig. 24—Voltage axial ratio vs number of turns, 13° helix; at upper frequency limits.

APPROXIMATE GRAPHICAL SOLUTION TO DETERMINANTAL EQUATION FOR MEDIUM-PITCH HELIX $\psi=13^\circ$

The solution to the determinantal equation for the infinite helix with $\psi=13^\circ$ has been given as curve (a) in Figs. 3 and 4. An alternative way of presenting the same information is to show the solution on the same graph as the pass band region, as illustrated in Fig. 25. This has an advantage in that it shows clearly the degree of approximation in an approximate graphical solution suggested by Sensiper,² which takes the form of the two straight lines AB and BC . AB lies along the edge of the forbidden region $m=-1$, and BC is that portion of the

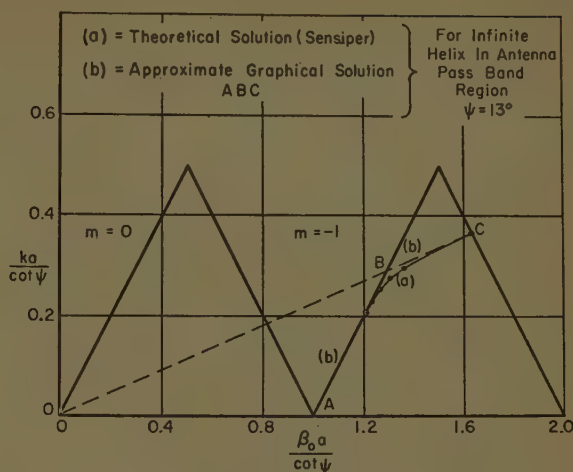


Fig. 25—Solutions of determinantal equation for $\psi=13^\circ$.

line $ka/\cot \psi = |\beta_0| a/\cot \psi \cdot \sin \psi$ that lies between the $m=-1$ and $m=-2$ forbidden regions.

It will be seen that the approximate graphical solution always gives a value of phase velocity along the axis of the antenna which is too high. Nevertheless, in the medium-pitch angle range it offers a quick method of estimating the upper frequency limit along BC , the solution obtained being about 15 per cent above the exact solution for long helices. It should be mentioned, perhaps, that this approximation involves only a few minutes work compared with the several days of computation required for the exact solution.

THE BANDWIDTH OF A NARROW-PITCH HELIX

An examination of the narrow-pitch helix was begun after the approximate graphical solution to the determinantal equation had indicated that propagation of the axial-beam mode with the proper phase velocity appeared to be possible even for narrow-pitch helices. It was further believed that the existence of this mode if, in fact, it did propagate, might have been overlooked in the past because of concentration on the center frequency of $C_\lambda=1$. The advantage of applying Sensiper's theory in this case is that it indicates that the first experimental check for the mode should be made in the vicinity of the lower frequency limit which is well below $C_\lambda=1$.

An experimental model was constructed with the smallest pitch angle which could be wound with enamelled wire of diameter 0.032 inch, at 3-cm wavelength—this was an angle of 1.8° . At this pitch angle, the turns were in physical contact and the helix, of length about 1.5λ , was self-supporting. No success was achieved with this model, and it was therefore decided to construct a longer one of length 4λ , wound with 181 turns. This was found to give good patterns and a good axial ratio, having an upper frequency limit of $0.81 C_\lambda$ compared with $1.1 C_\lambda$ for the 13° pitch helix. Three of the patterns obtained are shown in Figs. 26–28.

The propagation constant for this helix was solved

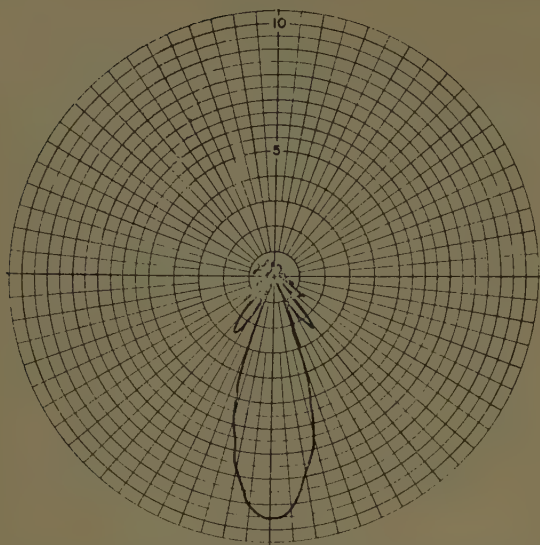


Fig. 26— E_θ pattern for 181-turn, 1.8° helix; $f=8$ kmc.

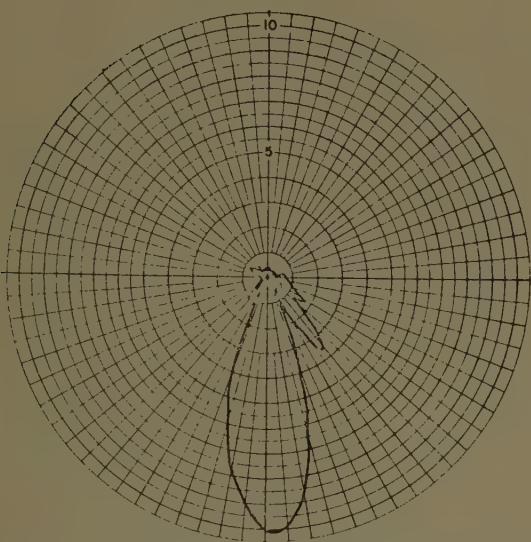


Fig. 27— E_ϕ pattern for 181-turn, 1.8° helix; $f=8$ kmc.

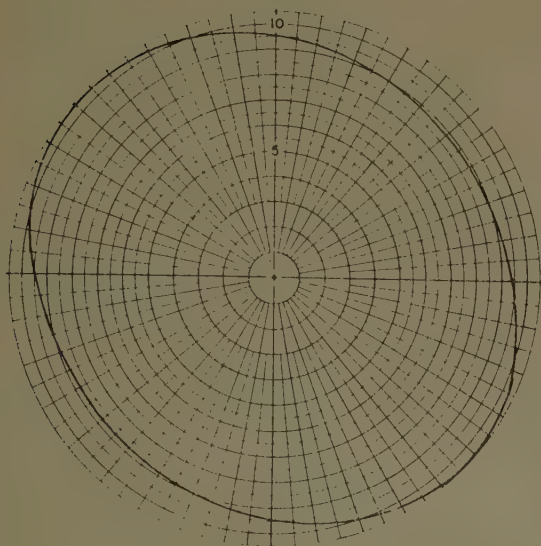


Fig. 28—Polarization pattern for 181-turn, 1.8° helix; $f=8$ kmc.

as in the medium-pitch case. One minor difference which should be noted is that asymptotic expressions for the modified Bessel functions must be used since the arguments of interest are outside the range of tabulated values. It was found that the upper frequency limit for the same criterion of sidelobe level as in the medium-pitch case was $0.87 C_\lambda$. The agreement with experimental results is therefore of the same order as for $\psi = 13^\circ$.

An additional antenna with the same pitch angle and of length 2λ was constructed and was found to give less satisfactory patterns, although it still satisfied the condition for increased directivity. It appears that it may be more difficult to excite the surface wave on a short narrow-pitch helix than on a medium-pitch helix of the same length. This may be because the axial phase velocity is less for $\psi = 1.8^\circ$ than for $\psi = 13^\circ$, so that a greater length of helix is necessary to bring about this reduction of velocity. Some evidence exists which suggests that the necessary over-all length may be reduced by starting off the helix winding with a medium-pitch turn before proceeding to the remainder of narrow-pitch. This will be discussed further in a separate report on the modulated helix.

No claim that the narrow-pitch-angle helix is superior to the medium-pitch helix in radiation pattern seems to be proper, nor is it expected that there will be any advantage in impedance characteristics. It is relevant to suggest, however, that for a constant length and constant diameter, there must be at least some difference in the axial phase velocity since it seems reasonable to presume that the different conductor dispositions result in different dielectric constants for the artificial dielectric medium. It would appear from past experiments that this effect is relatively small.

The elimination of this gap in the lower pitch-angle range is satisfying since, from a theoretical viewpoint, there is no valid reason for it to exist. It is now believed that a satisfactory pattern may be obtained for any pitch angle up to the value where the deterioration is caused by the pattern of the array factor.

CONCLUSIONS

It has been shown that the propagation constant determined from the theoretical solution for propagation along an infinite helix can be used to predict the upper frequency limit of a finite helical antenna of medium pitch angle, and of any length up to about ten wavelengths with an accuracy of the order of ten per cent.

This upper frequency limit decreases considerably with length, e.g., from $1.35 C_\lambda$ for a three-turn helix down to $1.0 C_\lambda$ for a 50-turn helix. The patterns in this range always satisfy the increased directivity condition, and the voltage axial ratios are less than 1.2 from 5 to 50 turns.

It has been established that radiation in the beam mode with good patterns and axial ratio can be obtained from a helical antenna four wavelengths long with a pitch angle as low as 1.8° .

Numerical Integration Methods for Antenna Pattern Calculations*

CHARLES C. ALLEN†

Summary—A study of numerical integration methods suitable for antenna pattern calculations was conducted for the purpose of determining which method provides a given accuracy with the greatest economy. The accuracy for a given method depends on the number of points at which the integrand is calculated, while the cost for a given number of points depends on the complexity of the method. This paper discusses the general principles of numerical integration, and outlines four methods in detail. The results of applying these methods on a digital computer to a simple cosine distribution are presented and analyzed. The relation of these results to those obtained for a pattern integral having a nonlinear phase function is discussed. The Gaussian quadratures are shown to have, in general, the highest degree of precision and lowest cost, while Filon's method is preferable when integrals having a linear phase function are calculated over a large range of pattern angle. The procedure for applying these methods to double numerical integration is outlined.

I. INTRODUCTION

THIS PAPER presents the results of a study of numerical integration methods suitable for antenna pattern calculations. While the examples use antenna pattern integrals, the methods discussed are general and may be applied to any problems in which an oscillating function is to be integrated. Typical fields in which such problems arise are acoustics, vibration, network theory, modulation, and astronomy. In many cases, the functions involved cannot be integrated literally and recourse is made to an approximation based on the first few terms of a series expansion. A more direct approach which preserves the original functions and their physical significance is to perform the integration numerically on a digital computer.

Complex problems which are being handled on the high-speed digital computers available today would have taken prohibitively long on the machines of a decade ago. With much shorter machine time now making such problems as the calculation of antenna patterns which require double integration of the current distribution on a shaped reflector feasible for computer calculation, closer attention must be paid to the efficiency of the methods of calculation used. While modern machines are fast, they are expensive to operate and calculation time must be kept to a minimum for economical operation.

The purpose of this study of methods of numerical integration was to determine their relative accuracy and costs. The accuracy for a given method depends on the number of points at which the integrand is calculated

while the cost for a given number of points depends on the complexity of the method. Only single numerical integration was investigated in this study since double numerical integration merely requires repeated application of the principles of single numerical integration as discussed in Section V of this paper. Further, if the integrand is a vector quantity and/or is complex, it can be resolved into separate vector components and/or real and imaginary components and each component integrated separately by numerical methods.

Four methods of numerical integration were studied, namely, the increment method, Simpson's rule, Filon's method, and Gaussian quadratures. The increment method is a simple one with which the author has obtained good results in calculating antenna patterns^{1,2} on some of the earlier computers. The results of this study show that Filon's method and Gaussian quadratures are much better suited for calculation of the more advanced antenna pattern integrals which present day computers are capable of handling.

A brief discussion of the methods of numerical integration which were studied is given in Section II. The results of applying these methods to a simple cosine distribution are presented in Section III. This distribution permits direct comparison with the exact solution obtained by literal integration and illustrates some characteristics of the different methods when applied to antenna pattern integrals. The results of applying two of the methods to an integrand having a nonlinear phase function are presented in Section IV. The latter integrand is representative of those encountered in calculating the scattering from objects or the patterns for non-collimating reflectors.

II. METHODS OF NUMERICAL INTEGRATION

A. General

The integrals encountered in the calculation of antenna patterns are generally of the form

$$E(u) = \int_a^b f(x) e^{j\Phi(u, x)} dx \quad (1)$$

where $f(x)$ is an amplitude function, $\Phi(u, x)$ is a phase function, and u is a parameter related to the pattern angle. Eq. (1) can be written in two parts as

¹ C. C. Allen, "Antenna pattern calculation for asymmetrical aperture distributions," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-3, p. 60-62; August, 1952.

² C. C. Allen, "Radiation patterns for aperture antennas with nonlinear phase distributions," 1953 IRE CONVENTION RECORD, pt. 2, pp. 9-17.

* The research reported in this paper has been sponsored by the Electronics Research Directorate of the AF Cambridge Res. Ctr., Air Res. and Dev. Command, under Contract AF 19(604)-1848.

† General Engrg. Lab., General Electric Company, Schenectady, N. Y.

$$p(u) = \int_a^b f(x) \cos \Phi(u, x) dx \quad (2)$$

and

$$q(u) = \int_a^b f(x) \sin \Phi(u, x) dx \quad (3)$$

such that $E(u) = p(u) + jq(u)$.

The integrands in (2) and (3) are seen to oscillate between $+f(x)$ and $-f(x)$ as an envelope with the number of oscillations depending on the change in $\Phi(u, x)$ within the interval of integration (a, b) . When $\Phi(u, x)$ is a linear phase function, such as ux , Filon's method can be applied directly to (2) and (3) as discussed in Section II-D. For other methods of numerical integration, however, the integrands in (2) and (3) will be considered to be the single functions $g_1(u, x)$ and $g_2(u, x)$, respectively.

Dropping the subscripts on $g(u, x)$ for purposes of general discussion and dropping u since it is held constant during integration, we are interested in studying numerical integration formulas of the form

$$\int_a^b g(x) dx \doteq \sum_{i=1}^n H_i g(x_i) \quad (4)$$

where the n values of $g(x_i)$ are ordinates or samples of the integrand and the n values of H_i are appropriate weight coefficients. If (4) is to be exact when $g(x)$ equals a constant, the sum of the n weight coefficients must equal $(b-a)$.

There are two general categories of such integration formulas which are discussed in considerable detail in recent texts^{3,4} on numerical analysis. The Newton-Cotes integration or quadrature formulas, which form one category, are based on Lagrangian methods and employ equally-spaced abscissae, x_i . The Gaussian quadrature formulas, which form the other category, employ abscissae, x_i , that are roots of orthogonal polynomials. Both categories are derived by representing the integrand, $g(x)$, by an approximating power polynomial.

According to the Weierstrass theorem on polynomial approximation,⁵ any continuous function can be approximated within the finite interval (a, b) as closely as desired by a power polynomial of sufficiently high degree. If none of the $n H_i$'s and the $n x_i$'s in (4) are arbitrarily assigned in advance, there will be $2n$ constants available which are sufficient to completely specify the definite integral of a power polynomial of degree $m = 2n - 1$. If the H_i 's and the x_i 's are properly chosen, therefore, (4) will be exact for $g(x)$ of degree m equal to $2n - 1$ or less. A quadrature formula having this degree of precision was derived by Gauss in a paper⁶ published in 1816 and is discussed in Section II-E.

If the $n x_i$'s are prescribed to be equally spaced, on the other hand, only n constants are available to specify the definite integral, and the degree of $g(x)$ for which (4) will be exact is reduced to m equal to $n - 1$ or less for even n and to n or less for odd n . The degree of precision of the Newton-Cotes formulas is, therefore, about one half that of the Gaussian quadratures.

When $g(x)$ is given as a table of equally-spaced ordinates, such as might be obtained from experimental measurements or a table of functions, one of the Newton-Cotes formulas is a logical choice for numerical integration. When $g(x)$ is known analytically, however, and the integration is to be performed on a digital computer, it is just as easy to calculate $g(x)$ at the irrational abscissae of a Gaussian quadrature formula as at the equally spaced abscissae of a Newton-Cotes formula. In applying a single integration formula to the interval (a, b) , therefore, only the Gaussian quadratures were investigated in this study.

The repeated application of simple two or three point Newton-Cotes formulas ($n = 2$ or 3) to equal subintervals of the interval (a, b) is often simpler than the use of a single quadrature formula employing the same total number of ordinates, on the other hand, and achieves satisfactory accuracy for many problems. The results are the familiar trapezoidal and parabolic or Simpson's rules,⁷ respectively, which are discussed in Sections II-B and II-C.

An extension of the Newton-Cotes formulas of the open type⁸ downward to a one-point formula not given in the referenced texts results in the increment formula of elementary calculus. Repeated application of the increment formula to equal subintervals of the interval (a, b) results in the increment method of numerical integration discussed in Section II-B.

Which integration formula should be used for a particular problem and the number of points required depends upon the nature of $g(x)$ and its derivatives. Expressions for the error incurred when a numerical integration formula of given precision is applied to a function $g(x)$ of higher degree are derived in the texts. Since these involve higher derivatives of $g(x)$, which are generally difficult to evaluate, their main usefulness is in comparing the order of magnitude of the error incurred by different formulas or the same formula with different n . The required derivatives are the n th for a Newton-Cotes formula with even n , the $(n+1)$ th for a Newton-Cotes formula with odd n , and the $2n$ th for a Gaussian quadrature.

Hildebrand has shown⁹ that for some functions, the repeated application of the two or three point Newton-Cotes formulas to equal subintervals can attain greater accuracy than the straightforward application of a Newton-Cotes formula having the same total number of

³ Z. Kopal, "Numerical Analysis," John Wiley & Sons, Inc., New York, N. Y., ch. VII; 1955.

⁴ F. B. Hildebrand, "Introduction to Numerical Analysis," McGraw-Hill Book Co., Inc., New York, N. Y., chs. 3 and 8; 1956.

⁵ Kopal, *op. cit.*, p. 19.

⁶ Kopal, *op. cit.*, p. 431.

⁷ Hildebrand, *op. cit.*, p. 75; Kopal, *op. cit.*, p. 406.

⁸ Hildebrand, *op. cit.*, p. 74.

⁹ Hildebrand, *op. cit.*, p. 81.

ordinates. In fact, the latter can actually result in an answer that diverges from the correct one after a certain value of n in this case. If the radius of convergence of the Taylor series expansion for $g(x)$ about all points in the interval (a, b) is greater than $(b-a)$, however, this difficulty does not occur.

B. Increment Method and Trapezoidal Rule

The increment method of numerical integration is a simple method which permits the integration process for a given function, $g(x)$, to be easily visualized. In applying this method, the interval (a, b) is divided into n equal subintervals or increments, $\Delta x = (b-a)/n$, and the ordinates are taken at the center of each increment. In each subinterval, the area under a graph of $g(x)$ is approximated by the rectangular area Δx times $g(x_i)$. The weight coefficients, H_i , are all equal to Δx , therefore, and (4) becomes

$$\int_a^b g(x) dx \doteq \Delta x \sum_{i=1}^n g(x_i). \quad (5)$$

Eq. (5) becomes exact in the limit as n approaches infinity for any $g(x)$ which is continuous in the interval (a, b) ; therefore, the integral can be calculated as accurately as desired by choosing n sufficiently large.

Eq. (5) is exact for n as small as one when $g(x)$ is of first degree [i.e., $g(x) = A_0 + A_1x$]. For $g(x)$ of higher degree, it is subject to an error of

$$E_I = -\frac{(b-a)^3}{24n^2} g''(x) = -\frac{(b-a)(\Delta x)^2}{24} g''(x) \quad (6)$$

where $g''(x)$ is the second derivative of $g(x)$ at some point x in the interval (a, b) . When $g(x)$ is of second degree, $g''(x)$ is a constant equal to $2A_2$ and E_I varies directly as $1/n^2$. For $g(x)$ of higher than second degree, $g''(x)$ depends on n and E_I varies as $g''(x)/n^2$.

For comparison with the increment method, consider the trapezoidal rule in which the interval (a, b) is divided into $(n-1)$ equal subintervals, $\Delta x = (b-a)/(n-1)$, and the n ordinates are taken at the ends of each subinterval. For the i th subinterval, the area under a graph of $g(x)$ is approximated by the trapezoidal area $\Delta x/2$ times $[g(x_i) + g(x_{i+1})]$. The resulting weight coefficients, H_i , in (4) are then $H_1 = \Delta x/2$, H_2 through $H_{n-1} = \Delta x$, and $H_n = \Delta x/2$.

The trapezoidal rule is exact for n as small as two when $g(x)$ is of first degree. For $g(x)$ of higher degree, it is subject to an error of

$$E_T = \frac{(b-a)^3}{12(n-1)^2} g''(x) = \frac{(b-a)(\Delta x)^2}{12} g''(x) \quad (7)$$

where $g''(x)$ is defined as before. Note that when $g(x)$ is of second degree, E_T has twice the magnitude of E_I and is of opposite sign for an equal number of subintervals.

For an equal number of ordinates, E_T is still larger. For $g(x)$ of higher than second degree, $g''(x)$ depends on n and will generally be larger than the value obtained with the increment method.

For the same number of ordinates, therefore, the increment method will generally have less than one half the error obtained with the trapezoidal rule. This can be seen intuitively by considering the error in the trapezoidal rule as the sum of the areas between arcs and chords of $g(x)$ and the error in the increment method as the sum of the areas between arcs and tangents of $g(x)$.

C. Simpson's Rule

With Simpson's rule for numerical integration, $g(x)$ is approximated with a series of short parabolic arcs rather than the tangents or chords of the increment method or trapezoidal rule, respectively. This, in general, reduces the error for a given number of ordinates since the area between $g(x)$ and its approximation is reduced in each subinterval.

In applying this rule, the interval (a, b) is divided into an even number of equal subintervals, $h = (b-a)/(n-1)$, and the odd number of n ordinates are taken at the ends of each subinterval. The Newton-Cotes three point formula is then applied to successive pairs of subintervals, $\Delta x = 2h$, so that the area under a graph of $g(x)$ is approximated in each double subinterval by $h/3$ times $[g(x_{\text{odd } i}) + 4g(x_{i+1}) + g(x_{i+2})]$. The resulting weight coefficients, H_i , in (4) are $H_1 = h/3$, H_2 through $H_{n-1} = 4h/3$ for even i , H_3 through $H_{n-2} = 2h/3$ for odd i , and $H_n = h/3$. Eq. (4) then becomes

$$\int_a^b g(x) dx \doteq \frac{h}{3} [g(x_1) + 4g(x_2) + 2g(x_3) + 4g(x_4) + \dots + 2g(x_{n-2}) + 4g(x_{n-1}) + g(x_n)] \quad (8)$$

where $x_1 = a$ and $x_n = b$.

Eq. (8) is exact for n as small as three when $g(x)$ is of third degree [i.e., $g(x) = A_0 + A_1x + A_2x^2 + A_3x^3$] or less. For $g(x)$ of higher degree, it is subject to an error of

$$E_S = \frac{(b-a)^5}{180(n-1)^4} g^{iv}(x) = \frac{(b-a)h^4}{180} g^{iv}(x) \quad (9)$$

where $g^{iv}(x)$ is the fourth derivative of $g(x)$ at some point x in the interval (a, b) . When $g(x)$ is of fourth degree, $g^{iv}(x)$ is a constant equal to $24A_4$ and E_S varies directly as $1/(n-1)^4$. For $g(x)$ of higher than fourth degree, $g^{iv}(x)$ depends on n and E_S varies as $g^{iv}(x)/(n-1)^4$.

The error obtained with Simpson's rule using n ordinates can be compared with the error obtained with the increment method using $(n-1)$ ordinates by substituting $(n-1)$ for n in (6). Δx in (6) is then equal to h in (9), and E_S is seen to be $-2h^2/15$ times E_I modified by the ratio of $g^{iv}(x_S)$ to $g''(x_I)$. This comparison is qualitative at best since even if the second and fourth derivatives of $g(x)$ were available, the proper values of x would be unknown.

D. Filon's Method

While numerical integration formulas corresponding to (4) can be derived for integrands of the form $f(x)w(x)$ where $w(x)$ is a weight function,¹⁰ the restrictions imposed on $w(x)$ do not permit the trigonometric weight functions in (2) and (3) to be used. Filon showed^{11,12} however, that when $\Phi(u, x)$ is a linear phase function ux , (2) and (3) can be solved directly by an extension of Simpson's rule. Filon's method is exact when $f(x)$ is of third degree or less, and for $f(x)$ of higher degree, the accuracy depends on how closely the integral of $f(x)$ alone can be approximated by Simpson's rule.

For this method, (2) and (3) become:

$$p(u) = \int_a^b f(x) \cos(ux) dx = \int_a^b g_1(u, x) dx \quad (10)$$

and

$$q(u) = \int_a^b f(x) \sin(ux) dx = \int_a^b g_2(u, x) dx. \quad (11)$$

As in Simpson's rule, the interval (a, b) is divided into an even number of equal subintervals, $h = (b-a)/(n-1)$, and the odd number of n ordinates are taken at the ends of each subinterval. The odd- and even-numbered ordinates are then summed separately for each equation as follows:

$$C_{\text{odd}} = g_1(u, a) + 2 \sum_{i=3}^{n-2} g_1(u, x_i) + g_1(u, b) \quad (\text{odd } i \text{ only})$$

$$C_{\text{even}} = 4 \sum_{i=2}^{n-1} g_1(u, x_i) \quad (\text{even } i \text{ only})$$

$$S_{\text{odd}} = g_2(u, a) + 2 \sum_{i=3}^{n-2} g_2(u, x_i) + g_2(u, b) \quad (\text{odd } i \text{ only})$$

$$S_{\text{even}} = 4 \sum_{i=2}^{n-1} g_2(u, x_i) \quad (\text{even } i \text{ only}).$$

Filon's numerical integration formulas for (10) and (11) are then:

$$p(u) \doteq h \{ \alpha(u) [g_2(u, b) - g_2(u, a)] + \beta(u) C_{\text{odd}} + \gamma(u) C_{\text{even}} \} \quad (12)$$

and

$$q(u) \doteq h \{ \alpha(u) [g_1(u, a) - g_1(u, b)] + \beta(u) S_{\text{odd}} + \gamma(u) S_{\text{even}} \} \quad (13)$$

where

$$\alpha(u) = \frac{(uh)^2 + (uh) \sin(uh) \cos(uh) - 2 \sin^2(uh)}{(uh)^3}$$

$$\beta(u) = \frac{(uh) + (uh) \cos^2(uh) - 2 \sin(uh) \cos(uh)}{(uh)^3}$$

$$\gamma(u) = \frac{\sin(uh) - (uh) \cos(uh)}{(uh)^3}.$$

When (uh) is small, the equations for $\alpha(u)$, $\beta(u)$, and $\gamma(u)$ can be solved more accurately on a digital computer by their series expansions. In the limit as u goes to zero, $\alpha(0)$ equals zero, $\beta(0)$ equals $\frac{1}{3}$, and $\gamma(0)$ equals $\frac{1}{3}$ in which case (12) becomes equal to (8) with $g(x)$ set equal to $f(x)$. The corresponding reduction of (13) to Simpson's rule is trivial since $g_2(u, x)$ is zero for all x when u equals zero.

E. Gaussian Quadratures

As indicated in Section II-A, the Gaussian quadratures have the highest degree of precision of all numerical integration formulas of the form given in (4). In fact, if $g(x)$ is of higher degree than $m = 2n - 1$, the error resulting from the use of an n point integration formula is a minimum for the Gaussian quadrature.¹³ The derivation of the Gaussian quadrature formula is treated thoroughly in the referenced texts and only the results will be presented here.

The n abscissae, x_i , are not equally spaced in the Gaussian quadrature, but are the roots of orthogonal polynomials which depend on the limits of integration. Likewise, the n weight coefficients, H_i , depend on the limits of integration since their sum must equal $(b-a)$. In order to prepare tables of x_i and H_i for different values of n , therefore, the interval (a, b) must be suitably normalized by a linear transformation. To do this, let the function to be integrated over the interval (a, b) be $G(s)$. Letting $s = (b+a)/2 + x(b-a)/2$, the integral becomes

$$\int_a^b G(s) ds = \frac{b-a}{2} \int_{-1}^1 g(x) dx \quad (14)$$

where $g(x) = G(s)$ for s corresponding to x . The quadrature formula for numerical integration over the normalized interval $(-1, 1)$ is then

$$\int_{-1}^1 g(x) dx \doteq \sum_{i=1}^n H_i g(x_i). \quad (15)$$

When (15) is a Gaussian quadrature, the n abscissae, x_i , are the roots of the Legendre polynomials, $P_n(x)$. The corresponding weight coefficients, H_i , are given by

$$H_i = -\frac{2}{(n+1)P_{n+1}(x_i)P_n'(x_i)}$$

$$= \frac{2}{nP_{n-1}(x_i)P_n'(x_i)}$$

¹³ Kopal, *op. cit.*, p. 359.

¹⁰ Kopal, *op. cit.*, p. 373.

¹¹ L. N. G. Filon, "On a quadrature formula for trigonometric integrals," *Proc. Roy. Soc. Edinburgh*, vol. 49, pp. 38-47; 1928.

¹² Kopal, *op. cit.*, p. 408.

where $P_n'(x_i)$ is the first derivative of $P_n(x)$ at x_i . The symmetry of the Legendre polynomials about the origin results in $x_i = -x_{n-i+1}$ and $H_i = H_{n-i+1}$. Tables of x_i and H_i for n equal to 2 through 16 are given by Kopal¹⁴ in his Appendix 4.1.

For comparison with the increment formula for a single subinterval, it is interesting to evaluate x_i and H_i for $n=1$. The results are $x_1=0$ and $H_1=2$ which correspond in un-normalized quantities to x_i at the center of the subinterval and H_i equal to Δx as given in Section II-B.

When $g(x)$ is of higher degree than $(2n-1)$, the Gaussian quadrature formula is subject to an error of

$$E_G = -\frac{2^{2n+1}(n!)^4}{(2n+1)[(2n)!]^3} g^{(2n)}(x) \\ = -\frac{(b-a)^{2n+1}(n!)^4}{(2n+1)[(2n)!]^3} G^{(2n)}(s) \quad (16)$$

where $g^{(2n)}(x)$ and $G^{(2n)}(s)$ are the $2n$ th derivatives of $g(x)$ and $G(s)$, respectively; at some point x in the interval $(-1, 1)$ and the corresponding point s in the interval (a, b) . Setting n equal to one reduces the second form of (16) to E_I for $n=1$ as given by (6).

III. PATTERN FOR COSINE DISTRIBUTION

A. True Pattern

The antenna pattern integral for a one-dimensional aperture distribution having uniform phase was used to investigate the accuracy of the methods of numerical integration discussed in Section II. This integral in physical units has the form

$$E(\phi) = (1 + \cos \phi) \int_{-a}^a F(s) e^{jks \sin \phi} ds \quad (17)$$

where $F(s)$ is the amplitude distribution, ϕ is the pattern angle measured from the normal to the aperture, $k=2\pi/\text{wavelength}$, and $2a$ is the length of the aperture. In this study, the term $(1 + \cos \phi)$ was dropped since it does not affect the integration.

It is customary to normalize the integral of (17) by letting $s=ax$ and $u=ka \sin \phi$. The normalized pattern integral is

$$E(u) = a \int_{-1}^1 f(x) e^{jux} dx. \quad (18)$$

For this study, the simple cosine distribution was selected; that is, $f(x) = \cos(\pi x/2)$. Since $f(x)$ is then an even function and $\sin ux$ is an odd function, $q(u)$ as defined by (3) is zero. The normalized pattern integral then becomes

$$E(u) = a \int_{-1}^1 \cos \frac{\pi x}{2} \cos ux dx \\ = 2a \int_0^1 \cos \frac{\pi x}{2} \cos ux dx. \quad (19)$$

The second form of (19) was used for the numerical integrations performed in this study.

Literal integration of (19) gives as the exact solution

$$E(u) = \frac{4a}{\pi} \frac{\cos u}{1 - \frac{4u^2}{\pi^2}}. \quad (20)$$

This is the true pattern with which the patterns obtained by numerical integration were compared. As is customary, the results of the exact and the numerical pattern calculations are presented in decibels below the peak of the main beam. The antenna pattern in decibels is given by

$$E(u)_{\text{db}} = 20 \log \frac{E(u)}{E(0)}. \quad (21)$$

The true pattern in decibels is, therefore,

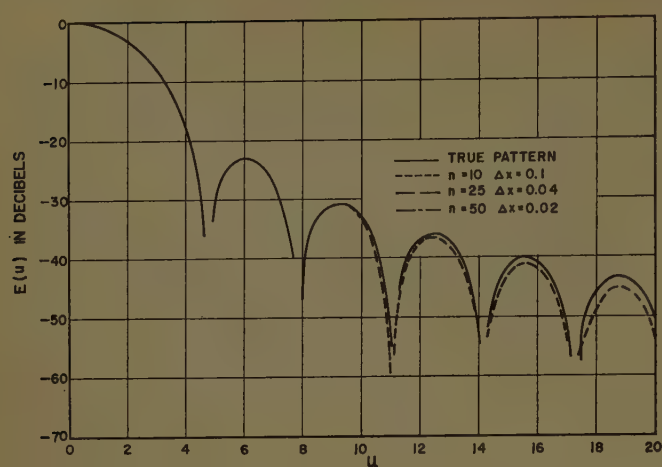
$$\text{True } E(u)_{\text{db}} = 20 \log \frac{\cos u}{1 - \frac{4u^2}{\pi^2}}. \quad (22)$$

The true pattern as given by (22) is shown by the solid lines in Fig. 1(a)–(e) for u from 0 to 100. This wide range of u was used in order to determine the accuracy of numerical integration when the integrand has many oscillations. The latter condition corresponds, for instance, to that encountered in scattering calculations. While more complicated integrals are generally involved in such calculations, they still have the basic form of (1).

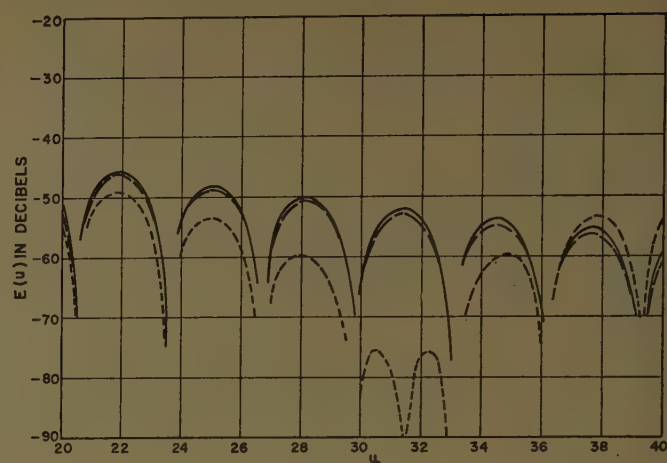
The numerical error in the value of an integral calculated by numerical methods is of primary interest from a mathematical point of view. For antenna pattern integrals, however, the error in decibels is generally of greater interest. The error incurred by applying a particular numerical integration formula to (19) is expressed in this paper as the difference between $E(u)_{\text{db}}$ obtained by numerical integration and the true $E(u)_{\text{db}}$. The error in decibels must be interpreted properly, however, since a small numerical error near a null of the pattern can result in a decibel error which is large in itself but small compared to the pattern level in decibels. What is of interest is the decibel error at the peaks of the sidelobes. For that reason, the decibel error is presented as the error in the envelope of the sidelobes wherever possible.

The numerical error corresponding to a given decibel error depends on the true numerical value to which the error applies. Table I shows for a range of decibel error the ratio of the numerical value of $E(u)$ to the true value of $E(u)$ and the corresponding numerical error relative to the true value of $E(u)$.

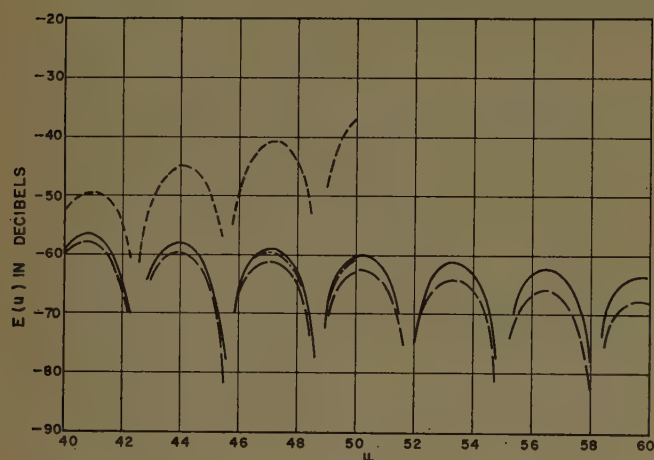
¹⁴ Kopal, *op. cit.*, p. 523.



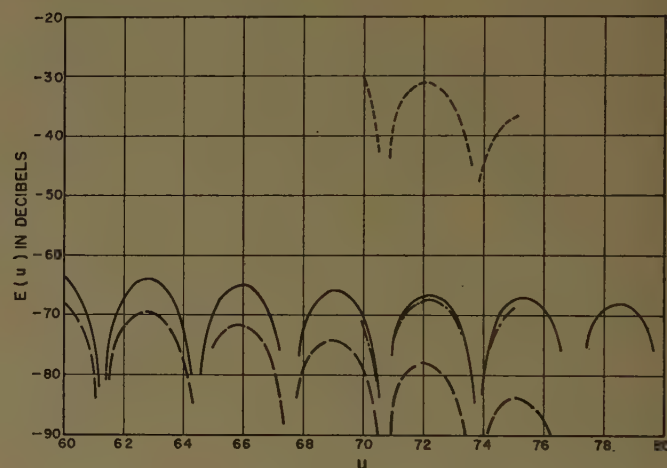
(a)



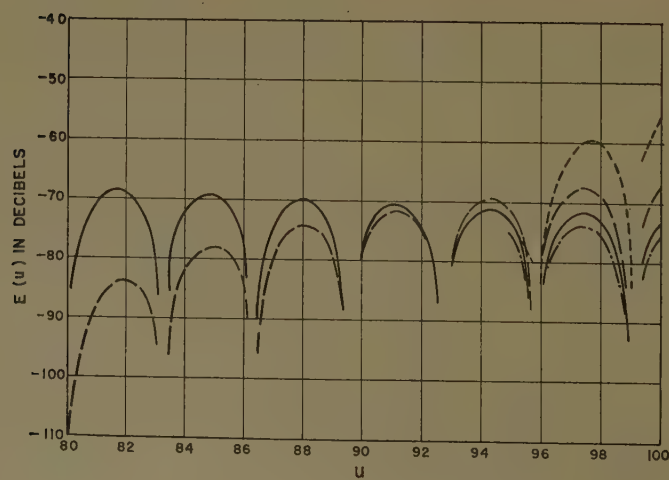
(b)



(c)



(d)



INCREMENT METHOD PATTERN (CONTINUED)

(e)

Fig. 1—(a) Antenna pattern for cosine distribution calculated by increment method. (b) Increment method pattern (continued). (c) Increment method pattern (continued). (d) Increment method pattern (continued). (e) Increment method pattern (continued).

TABLE I
NUMERICAL ERROR VS DB ERROR

| db Error | $R = \frac{\text{Numerical } E(u)}{\text{True } E(u)}$ | $(R-1) = \frac{\text{Numerical Error}}{\text{True } E(u)}$ |
|--------------|--|--|
| ± 20 | 10/0.1 | 9/-0.9 |
| ± 12 | 4/0.25 | 3/-0.75 |
| ± 6 | 2/0.5 | 1/-0.5 |
| ± 1 | 1.12/0.89 | 0.12/-0.11 |
| ± 0.1 | 1 ± 0.012 | ± 0.012 |
| ± 0.01 | 1 ± 0.0012 | ± 0.0012 |
| ± 0.001 | 1 ± 0.00012 | ± 0.00012 |
| ± 0.0001 | 1 ± 0.000012 | ± 0.000012 |

A larger decibel error is usually allowable at lower decibel levels since the corresponding numerical error is a smaller percentage of the maximum numerical value of the pattern. If a 1-db error exists at $E(u)_{ab} = -20$ db, for example, the numerical error is $0.12 E(u)$ or $0.012 E(0)$, while a 1-db error at $E(u)_{ab} = -60$ db results in a numerical error of $0.00012 E(0)$.

B. Pattern by Increment Method

The normalized pattern integral for a cosine distribution, (19), was calculated by the increment method discussed in Section II-B for n equal to 10, 25, and 50. The corresponding subintervals, Δx , are 0.1, 0.04, and 0.02. The patterns were calculated over the following ranges of u in steps of 0.5:

| n | Δx | Range of u | Δu |
|-----|------------|----------------------------|------------|
| 10 | 0.1 | 0-50, 70-75, 95-100 | 0.5 |
| 25 | 0.04 | 0-100 | 0.5 |
| 50 | 0.02 | 0-25, 45-50, 70-75, 95-100 | 0.5 |

The results of these calculations are plotted in Fig. 1(a)-(e) for comparison with the true pattern.

The decibel error in the envelope of the sidelobes for these calculations is shown in Fig. 2. For a given subinterval size, as u is increased to the value at which $u\Delta x = \pi$, the error builds up smoothly to a maximum value. The ordinate at the center of each subinterval is zero for this value of u and results in $E(u)$ obtained by numerical integration being equal to zero. For the cosine distribution, however, a sidelobe occurs at this value of u and results in the split lobe shown in Fig. 1 at $u = 31.4$ for $\Delta x = 0.1$. This condition occurs at $u = 78.5$ for $\Delta x = 0.04$.

Further increase in u to the value at which $u\Delta x = 2\pi$ (i.e., one cycle of ux per subinterval) results in each ordinate, $g(x_i)$, being equal to $-f(x_i)$. The value obtained for $E(u)_{ab}$ is then equal to $E(0)_{ab}$, and the decibel error to minus the true sidelobe level for that value of u .

Obviously, the increment method should not be used for u greater than some value smaller than that for which $u\Delta x = \pi$, the upper limit chosen depending upon the allowable error. Letting u_1 designate the value of u equal to $\pi/\Delta x$, the error curves shown in Fig. 2 form essentially a single curve when each is plotted against the renormalized pattern angle u/u_1 as shown in Fig. 3. This result indicates that the error for the increment

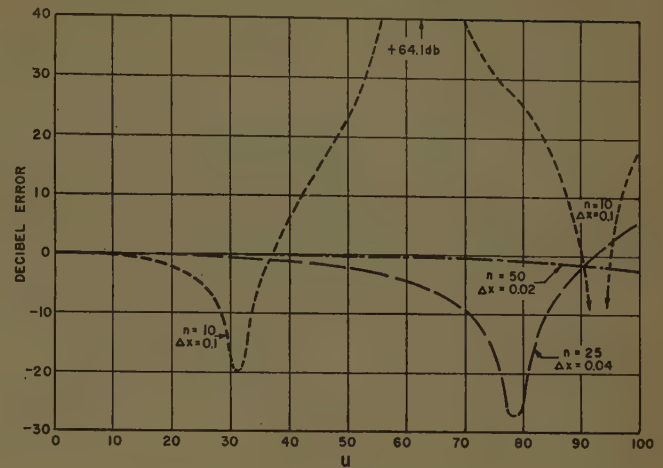


Fig. 2—Error in envelope of pattern for cosine distribution calculated by increment method.

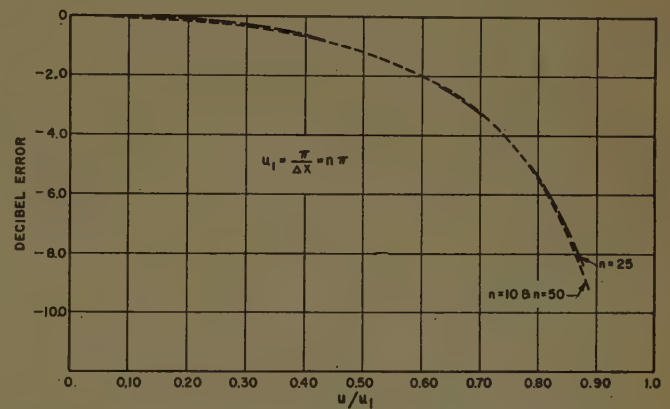


Fig. 3—Error in envelope of increment method pattern vs renormalized pattern angle, u/u_1 .

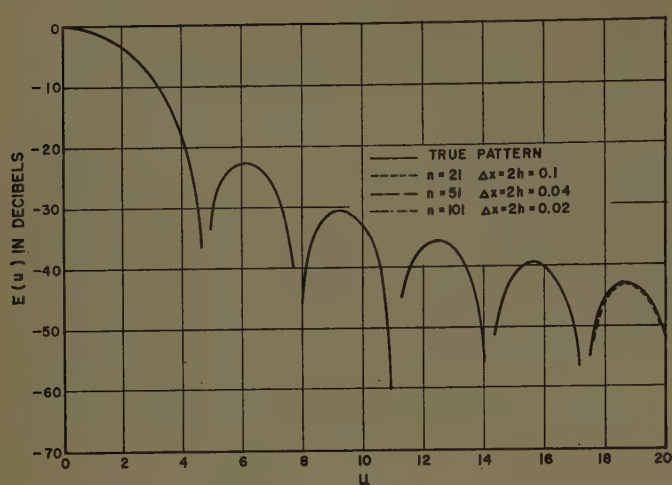
method is independent of the number of oscillations under $f(x)$, but depends instead upon the size of the subinterval relative to the period of oscillation.

As an example, Fig. 3 shows that the error will reach 1 db when u becomes equal to $0.47u_1 = 0.47\pi/\Delta x$. Conversely, to limit the error at u_{max} to 1 db, about two subintervals per half cycle of $u_{max}x$ are required. If four subintervals per half cycle of $u_{max}x$ are used instead, u_{max} will equal $0.25u_1$ and the error at $u_{max}x$ will be reduced to about 0.2 db. The application of these results to an integrand having a nonlinear phase function is discussed in Section IV.

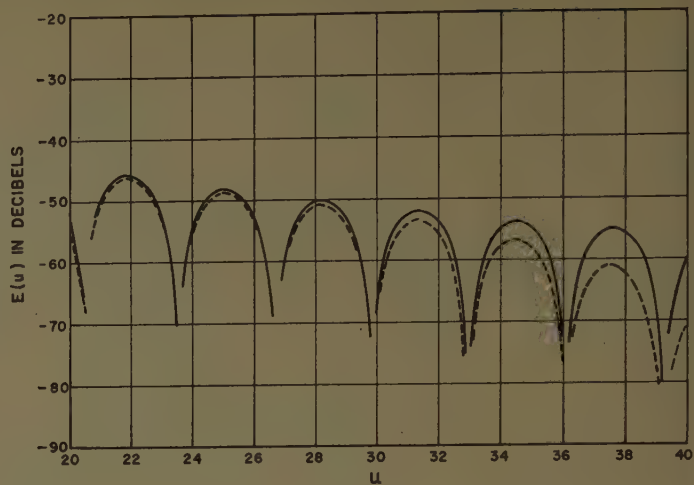
C. Pattern by Simpson's Rule

The normalized pattern integral for a cosine distribution, (19), was calculated by Simpson's rule discussed in Section II-C for n equal to 21, 51, and 101. The corresponding double subintervals, $\Delta x = 2h$, are 0.1, 0.04, and 0.02. The patterns were calculated over the following ranges of u in steps of 0.5:

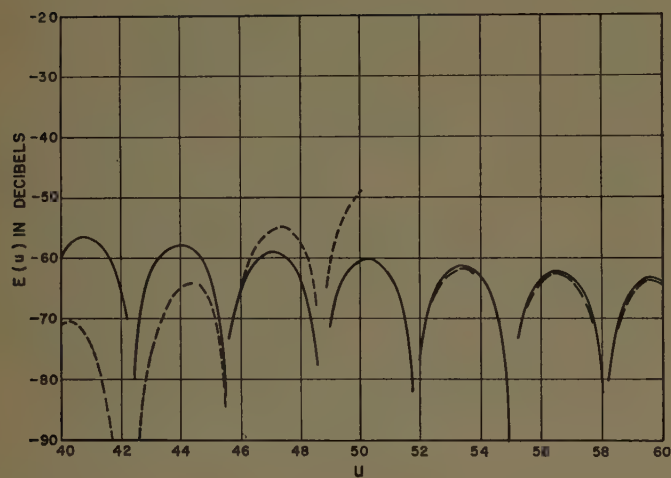
| n | $\Delta x = 2h$ | Range of u | Δu |
|-----|-----------------|----------------------------|------------|
| 21 | 0.1 | 0-50, 70-75, 95-100 | 0.5 |
| 51 | 0.04 | 0-100 | 0.5 |
| 101 | 0.02 | 0-25, 45-50, 70-75, 95-100 | 0.5 |



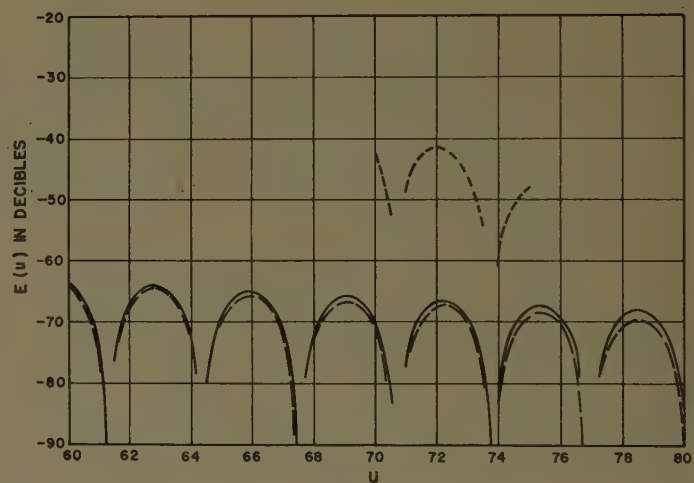
(a)



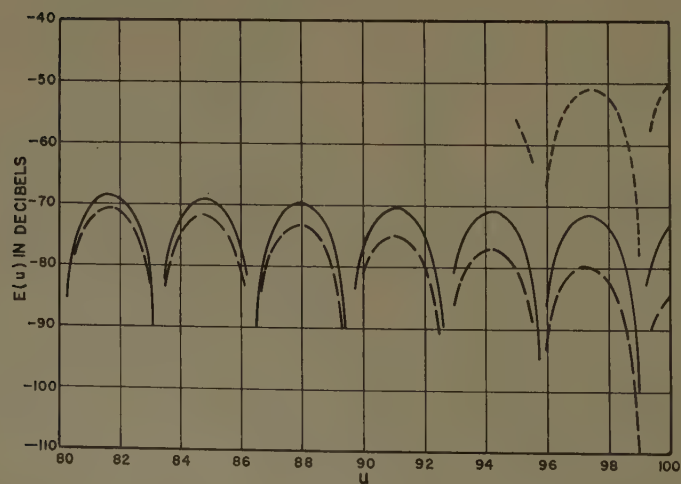
(b)



(c)



(d)



(e)

Fig. 4—(a) Antenna pattern for cosine distribution calculated by Simpson's rule. (b) Simpson's rule pattern (continued). (c) Simpson's rule pattern (continued). (d) Simpson's rule pattern (continued). (e) Simpson's rule pattern (continued).

The results of these calculations are plotted in Fig. 4(a)-(e) together with the true pattern for comparison.

The decibel error in the envelope of the sidelobes for these calculations is shown in Fig. 5. The error for Simpson's rule changes with increasing u in a more complex manner than the error for the increment method. For a double subinterval, $\Delta x = 2h$, equal in size to a subinterval of the increment method, the error for Simpson's rule is seen to increase more slowly at first than the error for the increment method. Once u becomes larger than $u_1 = \pi/\Delta x$, however, the error for Simpson's rule increases rapidly to a maximum at about $u = 1.3u_1$ after which it decreases rapidly, changes sign, and reaches a maximum near $u = 2u_1$ which is equal to about 10 db less than minus the sidelobe level at $2u_1$. At $u = 2u_1$, the odd ordinates equal $+f(x_i)$ while the even ordinates equal $-f(x_i)$ which results in a value of $E(2u_1)$ equal to about $-E(0)/3$. Finally, at $u = 4u_1$ there is one full cycle per subinterval, h , and all ordinates equal $+f(x_i)$ with the result that $E(4u_1)$ equals $E(0)$ and the decibel error equals minus the true sidelobe level.

As with the increment method, therefore, Simpson's rule should not be used for u greater than some value smaller than u_1 , the value of u_{\max} depending on the allowable error. If the decibel error curves for Simpson's rule were plotted against the renormalized pattern angle u/u_1 , as was done for the increment method, it is seen from Fig. 5 that essentially a single error curve would be obtained. This indicates that the error for Simpson's rule depends mainly upon the size of the subinterval relative to the period of oscillation as was observed for the increment method.

Simpson's rule results in a 1-db error at about twice the renormalized angle u/u_1 for which a 1-db error is obtained with the increment method. For the same number of ordinates, u_1 would be about half as large for Simpson's rule as for the increment method, however, and the values of u at which 1-db error is obtained would be about the same. For a given error of less than one decibel, on the other hand, the value of u for Simpson's rule is greater than for the increment method with the same number of ordinates. This can be observed by comparing the curve for $n = 51$ in Fig. 5 with the curve for $n = 50$ in Fig. 2.

D. Pattern by Filon's Method

The normalized pattern integral for a cosine distribution, (19), was calculated by Filon's method discussed in Section II-D for n equal to 9 and 21. The corresponding double subintervals, $\Delta x = 2h$, are 0.25 and 0.1. These patterns were calculated over the range of u from 0 to 100 in steps of 0.5. Most of the sidelobes for the calculated patterns were within 0.1 db of the true pattern and none differed by more than 0.4 db.

The decibel error in the envelope of the sidelobes for these calculations is shown in Fig. 6 with the error at each of the sidelobe maxima indicated by a small circle. The error is seen to behave in a systematic manner

about values of u which are multiples of $u_2 = 2\pi/\Delta x = 2u_1$. These values of u correspond to the positions of the $[M(n-1)-1]$ th sidelobes for integer values of M . When $u = Mu_2$, there are M complete cycles of ux in each double subinterval with the result that the positions of the nulls on either side of Mu_2 are sensitive to the departure from $f(x)$ of its parabolic approximations. Because of the rapid changes in these regions, the actual error has been shown rather than the error in the envelope of the sidelobes.

As u is increased toward u_2 , the decibel error goes to minus infinity at the value of u just before u_2 for which $E(u)$ calculated by this method is zero. The decibel error then changes rapidly through zero to plus infinity at the value of u for which the true $E(u)$ is zero (i.e., $u = u_2 - \pi/2$); however, the separation in u between these two null positions is so small that the nearly vertical line is not shown in Fig. 6. From plus infinity, the decibel error decreases to zero near u_2 and continues on to minus infinity at the value of u just after u_2 for which $E(u)$ calculated by this method is zero. Finally, the

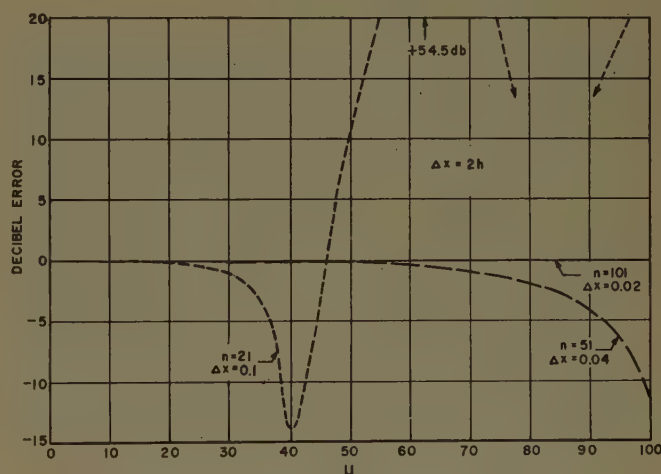


Fig. 5—Error in envelope of pattern for cosine distribution calculated by Simpson's rule.

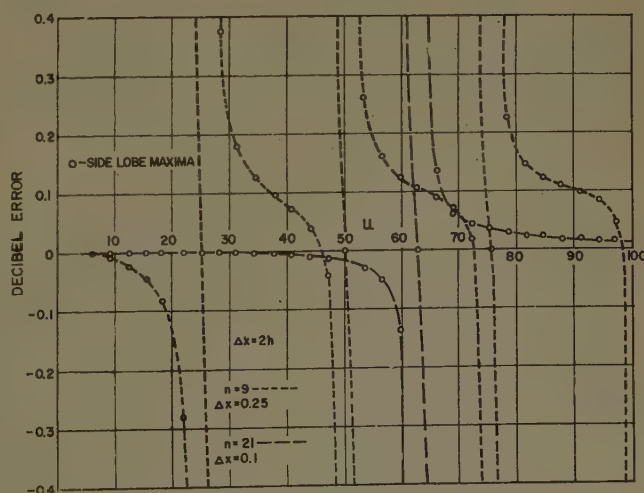


Fig. 6—Error in envelope of pattern for cosine distribution calculated by Filon's method.

decibel error changes rapidly through zero to plus infinity at $u = u_2 + \pi/2$ (line not shown in Fig. 6) after which the error decreases smoothly for several sidelobes until the next multiple of u_2 is approached. The behavior just described is then repeated.

The large changes in the decibel error in the regions near $u = Mu_2$ will generally correspond to small numerical errors as discussed in Section III-A since they occur near the pattern nulls and result from the calculated nulls being slightly displaced from the true nulls. In general, all of the pattern nulls calculated by numerical methods are displaced to some extent from the true pattern nulls which results in the decibel error between all sidelobes behaving in a manner similar to that discussed above. These effects are less pronounced than in the regions near $u = Mu_2$, however, since reinforcement of the effect from each subinterval does not occur. In this calculation, for example, the only significant deviations from the curve for $n = 21$ plotted in Fig. 6 occurred at values of u equal to 11, 33, 55, 77, and 99 which are very near the nulls of the true pattern. The corresponding deviations are only 0.02, 0.06, 0.03, 0.02, and 0.01 db.

E. Pattern by Gaussian Quadratures

The normalized pattern integral for a cosine distribution, (19), was calculated by Gaussian quadratures, discussed in Section II-E, for n equal to 5 and 10. The patterns were calculated over the range of u from 0 to 50 in steps of 0.5. Since the table of abscissae and weight coefficients referenced in Section II-E is based on the normalized interval $(-1, 1)$, the proper abscissae for use with the interval $(0, 1)$ of (19) were obtained by letting $x_i = 0.5 + 0.5X_i$, where X_i are the abscissae obtained from the table. The corresponding weight coefficients for use with the interval $(0, 1)$ are one half the H_i obtained from the table. This corresponds to the transformation indicated in (14).

The decibel error for these calculations is shown in Fig. 7 with the error indicated by a small circle at each of the sidelobe maxima and by a small + at each of the nulls. As might be expected for a formula having a high degree of precision, the decibel error remains very close to zero as u is increased over a useful range from 0 to some value u_{\max} beyond which the error behaves in an erratic manner and becomes quite large. Since the ab-

scissae are not equally spaced, the error does not build up systematically at particular values of u as it does for the methods previously discussed. In general, large changes in decibel error occur for u greater than u_{\max} as a result of the calculated nulls being displaced from the true pattern nulls as discussed in Section III-D.

For $n = 5$, u_{\max} is about 8 which corresponds closely to the degree of precision of the 5 point Gaussian quadrature formula, $m = 2n - 1 = 9$. Likewise for $n = 10$, u_{\max} is about 21 which corresponds closely to the degree of precision $m = 19$. This correspondence can be analyzed in terms of the degree required for a power polynomial to closely approximate $g(u_{\max}, x)$. Since there are u_{\max}/π half cycles in the interval $(0, 1)$, the polynomial must have at least u_{\max}/π roots which places a lower limit on the required degree. Assuming that a polynomial of degree equal to about three times the number of roots in the interval $(0, 1)$ will adequately approximate $g(u_{\max}, x)$, the required degree of precision is about equal to u_{\max} . The required number of ordinates, n , is then equal to about $u_{\max}/2$, which agrees with the results noted above.

F. Comparison of Results

The results of the calculations discussed in Sections III-B through E are compared in Table II which shows the upper value of u for a given decibel error over the

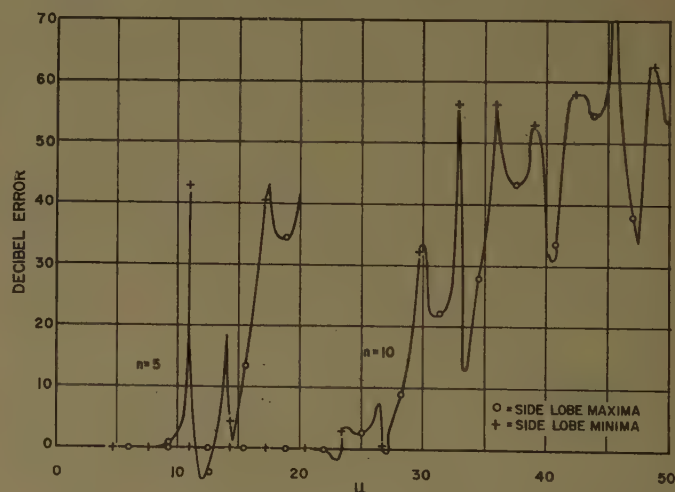


Fig. 7—Error in envelope of pattern for cosine distribution calculated by Gaussian quadratures.

TABLE II
UPPER VALUE OF u FOR GIVEN DB ERROR

| db Error Less Than | Increment Method | | | Simpson's Rule | | | Filon's Method | | Gaussian Quadrature | |
|-----------------------|------------------|----------|------|----------------|----------|------|----------------|----------|---------------------|----------|
| | 10 | $n = 25$ | 50 | 21 | $n = 51$ | 101 | 9 | $n = 21$ | 5 | $n = 10$ |
| 0.0001 | 0 | 0 | 0.5 | 3.5 | 3.0 | 3.0 | 1.0 | 4.0 | 2.5 | 2.0 |
| 0.001 | 0 | 2.0 | 2.5 | 4.0 | 13.5 | 22.5 | 2.5 | 10.5 | 4.0 | 4.0 |
| 0.01 | 2.0 | 4.5 | 7.5 | 10.5 | 26.5 | 48.5 | 8.5 | 43.5 | 5.5 | 16.5 |
| 0.1 | 5.0 | 13.0 | 25.0 | 17.0 | 43.5 | >75 | 19.0 | 58.5 | 8.0 | 21.0 |
| 1.0 | 14.5 | 36.5 | 73.5 | 28.5 | 71.0 | >100 | 23.0* | 61.0* | 9.5 | 22.5 |

* See text below.

useful range of allowable error. The most meaningful comparisons are between the increment method for $n=50$ and Simpson's rule for $n=51$, between Simpson's rule for $n=21$ and Filon's method for $n=21$, and between the increment method for $n=10$, Filon's method for $n=9$, and the Gaussian quadrature for $n=10$. The last comparison clearly demonstrates the high degree of precision of the Gaussian quadrature.

The values of u_{\max} shown in Table II for a 1.0-db error with Filon's method are somewhat artificial since they result from the behavior of the decibel error curves near $u=u_2$ as discussed in Section III-D. Disregarding the error in this region, which is actually well under 1.0 db at the sidelobe maxima, the upper limit of u for a 1.0-db error was not reached for this method. In fact, the errors at the sidelobe maxima near $u = Mu_2$ appear to be decreasing with increasing M .

The time per ordinate per step in u required to calculate the pattern integral on a digital computer was found to be about the same for all four methods except for some additional time per step in u required to calculate $\alpha(u)$, $\beta(u)$, and $\gamma(u)$ in Filon's method. This additional time was about 56 per cent for $n=9$ and about 24 per cent for $n=21$. For the same range and number of steps in u , therefore, the relative costs of calculating a pattern by each of these methods are nearly in proportion to the number of ordinates required for each to obtain the desired accuracy with the exception that the cost for Filon's method is proportional to n increased by a factor that depends on n .

IV. PATTERN FOR NONLINEAR PHASE FUNCTION

A. Pattern Integral

The antenna pattern integral for an integrand having a nonlinear phase function was calculated by the increment method and by Gaussian quadratures in order to study the application of the results of Section III to such integrals. The geometry for the integral studied is shown in Fig. 8 and represents a cylindrical reflector illuminated by a point source located at F . The quantities shown in Fig. 8 result in an integrand having a large number of oscillations between the limits of integration.

The single integral of the current distribution on the reflector was taken in the x - y plane with limits of integration equal to the angles $\pm a_m$ at which the feed vector $\bar{\rho}$ is tangent to the reflector. With the phase reference at F , the pattern integral is

$$E(\phi) = \int_{-a_m}^{a_m} \frac{H(\gamma)}{\rho(a)} P(a, \phi) e^{-jk\rho(a)[1+\cos(\phi+\gamma)]} R da \quad (23)$$

where γ is a function of a , $H(\gamma)$ is the field intensity pattern of the feed in the x - y plane, $k=2\pi/\text{wavelength}$, and $P(a, \phi)$ results from the vector product $\bar{n} \times (\bar{\rho}_1 \times \bar{e})$ which appears in the integral of the current distribution on a reflector.¹⁵

¹⁵ S. Silver, "Microwave Antenna Theory and Design," McGraw-Hill Book Co., Inc., New York, N. Y., p. 150; 1949.

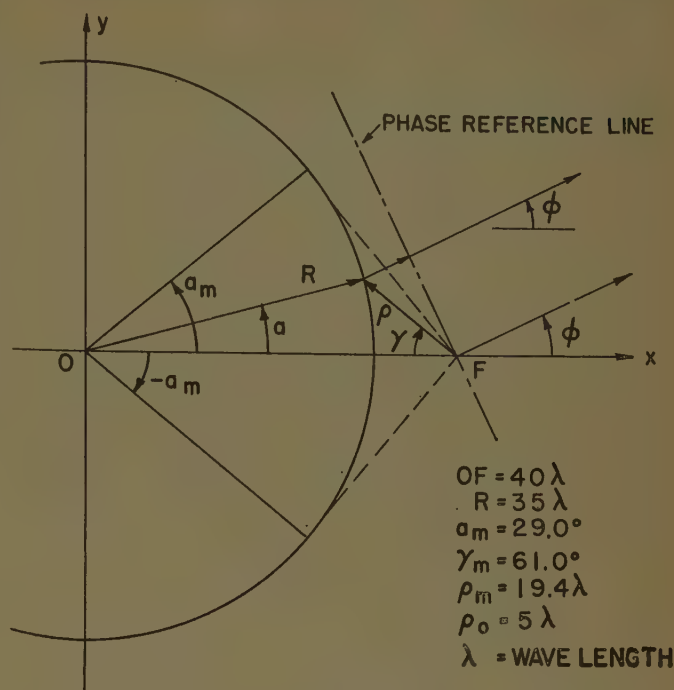


Fig. 8—Geometry for pattern integral with nonlinear phase function.

For a feed polarized in the z direction (perpendicular to the x - y plane), $P(a, \phi) = \cos(a + \gamma)$ and is independent of ϕ . For a feed polarized in the x - y plane and perpendicular to $\bar{\rho}$, $P(a, \phi) = \cos(a - \phi)$. Since the polarization is not pertinent to the present study of methods of numerical integration, $P(a, \phi)$ was taken as unity in the pattern calculated. This simplification merely results in an amplitude function that is greater than that for either polarization over most of the interval of integration.

Letting $E(\phi) = p(\phi) + jq(\phi)$ for purposes of numerical integration, (23) assumes the form of (2) and (3):

$$p(\phi) = R \int_{-a_m}^{a_m} f(a) \cos \Phi(a, \phi) da \quad (24)$$

and

$$q(\phi) = -R \int_{-a_m}^{a_m} f(a) \sin \Phi(a, \phi) da \quad (25)$$

where

$$f(a) = \frac{H(\gamma)}{\rho(a)},$$

and

$$\Phi(a, \phi) = k\rho(a)[1 + \cos(\phi + \gamma)].$$

The amplitude function, $f(a)$, for $H(\gamma) = \cos^2 \gamma$ is shown in Fig. 9 for the dimensions given in Fig. 8. The corresponding phase functions, $\Phi(a, \phi)$, for $\phi = 0^\circ$, 30° , and 60° are shown in wavelengths in Fig. 9. In the calculations discussed in Sections IV-B and C, the constant R in (24) and (25) has been dropped.

B. Pattern by Increment Method

The increment method discussed in Section II-B was used to calculate the integrals in (24) and (25) for the functions shown in Fig. 9. The interval $(-29.0^\circ, 29.0^\circ)$ was divided into 580 equal subintervals, $\Delta a = 0.1^\circ$, in order to obtain fairly accurate integrations with which the results of using larger increments and Gaussian quadratures could be compared. For $n = 580$, the minimum number of subintervals per half cycle of $\Phi(a, \phi)$ for $\phi = 0^\circ, 30^\circ$, and 60° is 5.5, 4.4, and 4.1, respectively, near $a = -29^\circ$. The number of subintervals per half cycle is much greater in the region of stationary phase (*i.e.*, near zero slope of the phase function) which provides the major contribution to the integral. There are, for example, about 46 subintervals between $a = -2.3^\circ$ and $+2.3^\circ$ which provide the major contribution to $p(0^\circ)$.

The results of these calculations are:

Interval $(-29.0^\circ, 29.0^\circ)$, $\Delta a = 0.1^\circ$, $n = 580$

$$\begin{aligned} E(0^\circ) &= 0.464\ 727 - j0.441\ 472 \text{ or, } -0.0000 \text{ db} \\ E(30^\circ) &= 0.527\ 252 + j0.219\ 958 \text{ or, } -0.9999 \text{ db} \\ E(60^\circ) &= 0.141\ 333 + j0.368\ 132 \text{ or, } -4.2198 \text{ db} \end{aligned}$$

All calculations were made using eight significant figures, but the results have been rounded off to six. $E(0^\circ)$ for $\Delta a = 0.1^\circ$ has been taken as the zero decibel level for purposes of comparison.

To better visualize the integration process, $p(\phi, a)$ and $q(\phi, a)$ have been plotted for $\phi = 0^\circ, 30^\circ$, and 60° as a function of the upper limit, a , as it is increased from -29.0° to $+29.0^\circ$. Subtotals of the numerical integra-

tion were printed out by the digital computer every 0.2° in a to provide these integral functions which are shown in Figs. 10–15. The oscillations of the integral are $\frac{1}{4}$ cycle out of phase with the oscillations of the integrand since as a positive half cycle of the integrand is completed, the integral must reach a positive maximum and then decrease as the integrand begins a negative half cycle. As the stationary phase region is approached, the amplitude of oscillation of the integral increases due principally to the increase in the period of oscillation and also to the

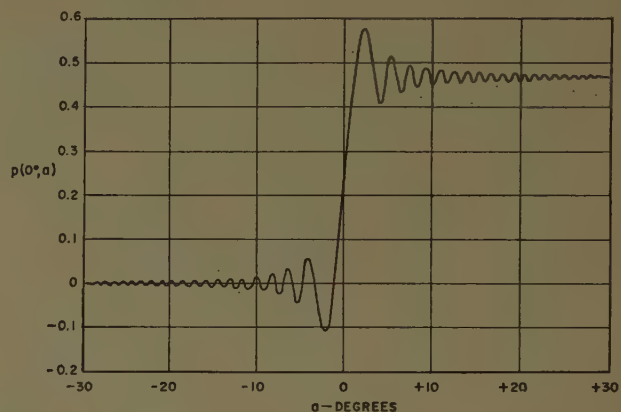


Fig. 10—Integral $p(\phi, a)$ for nonlinear phase function with $\phi = 0^\circ$.

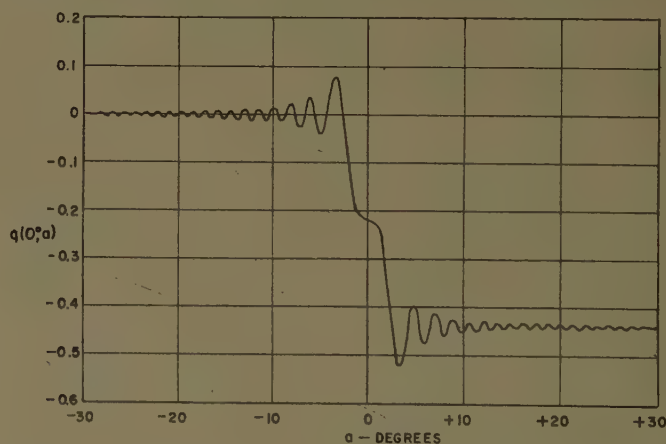


Fig. 11—Integral $q(\phi, a)$ for nonlinear phase function with $\phi = 0^\circ$.

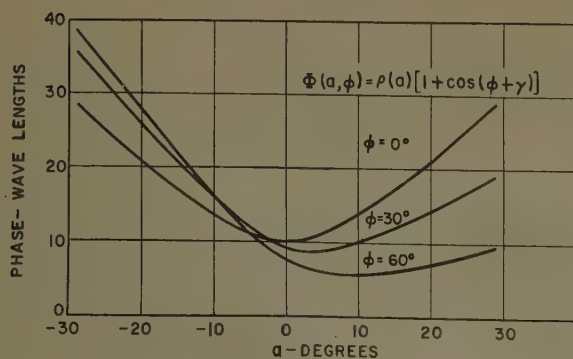
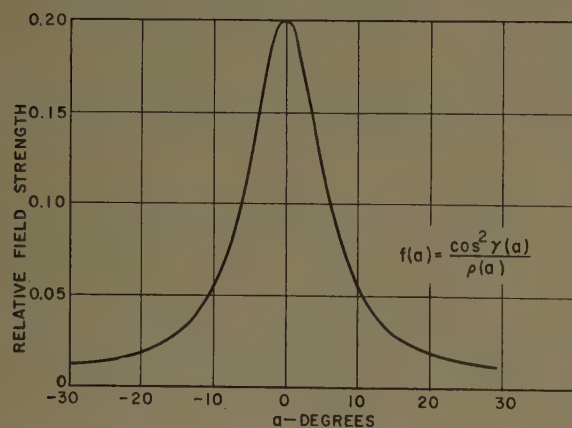


Fig. 9—Amplitude and phase functions for $\phi = 0^\circ, 30^\circ$, and 60° .

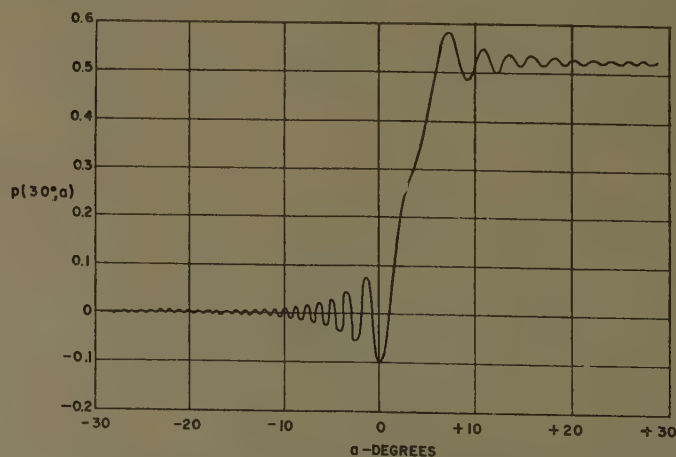


Fig. 12—Integral $p(\phi, a)$ for nonlinear phase function with $\phi = 30^\circ$.

increase in the amplitude function, $f(a)$. After the major contribution to the integral occurs in the stationary phase region, the amplitude of oscillation of the integral decreases as the period of oscillation decreases.

The results of numerical integration using increments equal to 0.3° were obtained at the same time that the calculations for $\Delta a = 0.1^\circ$ were done by summing separately every third ordinate starting with the second (*i.e.*, the middle of the first 0.3° subinterval). This reduced the number of subintervals per half cycle to $\frac{1}{3}$ of that for $\Delta a = 0.1^\circ$. Since 58.0° is not a multiple of 0.3° , the results of numerical integration for the two sub-

interval sizes are compared below for the interval $(-29.0^\circ, 28.6^\circ)$.

Interval $(-29.0^\circ, 28.6^\circ)$, $\Delta a = 0.1^\circ/0.3^\circ$, $n = 576/192$

$$E(0^\circ) = 0.467\ 923 - j0.443\ 767 \text{ or, } +0.0527 \text{ db} \\ 0.467\ 798 - j0.443\ 940 \text{ or, } +0.0530 \text{ db}$$

$$E(30^\circ) = 0.523\ 239 + j0.218\ 170 \text{ or, } -1.0669 \text{ db} \\ 0.523\ 374 + j0.217\ 785 \text{ or, } -1.0672 \text{ db}$$

$$E(60^\circ) = 0.143\ 691 + j0.372\ 269 \text{ or, } -4.1168 \text{ db} \\ 0.143\ 403 + j0.372\ 331 \text{ or, } -4.1178 \text{ db}$$

In relating these results to those discussed in Section III-B for a linear phase function, the maximum slope of $\Phi(a, \phi)$ in radians per unit of a corresponds to the value of u_{\max} discussed in connection with Fig. 3, and a slope equal to $\pi/\Delta a$ corresponds to u_1 . The maximum slopes of the phase functions shown in Fig. 9 are ± 5.70 , -7.10 , and -7.66 radians per degree of a for $\phi = 0^\circ$, 30° , and 60° , respectively. The corresponding values of u/u_1 for $\Delta a = 0.1^\circ$ are 0.181, 0.226, and 0.244 which, according to Fig. 3, would result in errors of 0.1 to 0.2 db for linear phase functions having these slopes. The error resulting from the use of $\Delta a = 0.1^\circ$ is considerably less than 0.1 db, however, since the effective values of u/u_1 are much less than the maximum values given above. The low effective values of u/u_1 result from the small slope of $\Phi(a, \phi)$ in the region of stationary phase and the progressively changing relation between the size of Δa and the period of oscillation of $\Phi(a, \phi)$. The close agreement between the results for $\Delta a = 0.3^\circ$ and 0.1° indicate that these two factors are very effective in keeping the error small.

C. Pattern by Gaussian Quadratures

The Gaussian quadratures discussed in Section II-E were used to calculate the integrals in (24) and (25) for the functions shown in Fig. 9. Because of the large number of oscillations of the integrands, a single quadrature formula was not used. Instead, the Gaussian quadrature formula for $n = 16$ was applied to each of 10 equal subintervals of the interval $(-29.0^\circ, 29.0^\circ)$ for $\phi = 0^\circ$, 30° , and 60° . The corresponding total number of ordinates is $n_T = 160$. The 16 point formula was also applied to each of 6, 9, and 12 equal subintervals of the interval $(-29.0^\circ, 29.0^\circ)$ for $\phi = 60^\circ$, the corresponding values of n_T being 96, 144, and 192. The results of these calculations are compared below with the results of the increment method for $n = 580$.

| n_T | $p(0^\circ)$ | $+j$ | $q(0^\circ)$ | $E(0^\circ)$ db |
|-------|---------------|----------------|---------------|------------------|
| 160 | 0.464 784 | $-j0.441\ 483$ | or, | $+0.0007$ db |
| 580 | 0.464 727 | $-j0.441\ 472$ | or, | 0.0000 db |
| n_T | $p(30^\circ)$ | $+j$ | $q(30^\circ)$ | $E(30^\circ)$ db |
| 160 | 0.527 247 | $+j0.219\ 903$ | or, | -1.0003 db |
| 580 | 0.527 252 | $+j0.219\ 958$ | or, | -0.9999 db |
| n_T | $p(60^\circ)$ | $+j$ | $q(60^\circ)$ | $E(60^\circ)$ db |
| 96 | 0.133 527 | $+j0.436\ 480$ | or, | -2.9493 db |
| 144 | 0.141 062 | $+j0.368\ 046$ | or, | -4.2238 db |
| 160 | 0.141 371 | $+j0.368\ 198$ | or, | -4.2182 db |
| 192 | 0.141 366 | $+j0.368\ 131$ | or, | -4.2196 db |
| 580 | 0.141 333 | $+j0.368\ 132$ | or, | -4.2198 db |

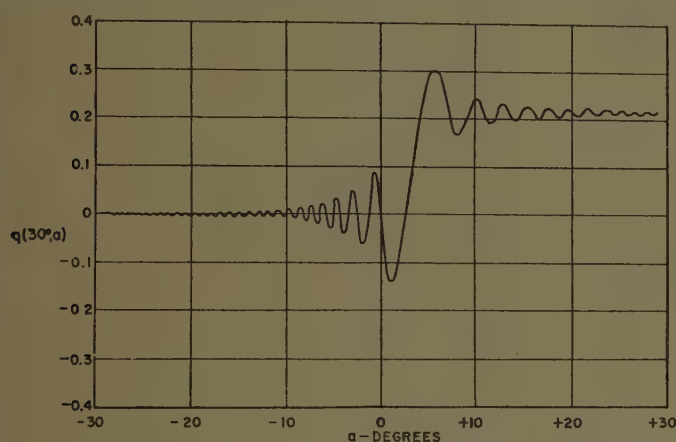


Fig. 13—Integral $q(\phi, a)$ for nonlinear phase function with $\phi = 30^\circ$.

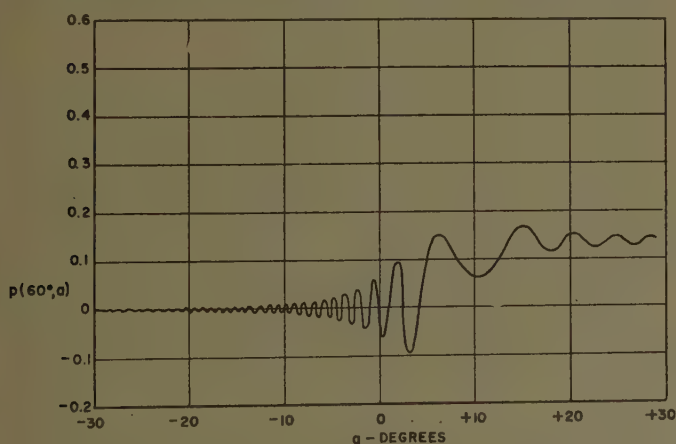


Fig. 14—Integral $p(\phi, a)$ for nonlinear phase function with $\phi = 60^\circ$.

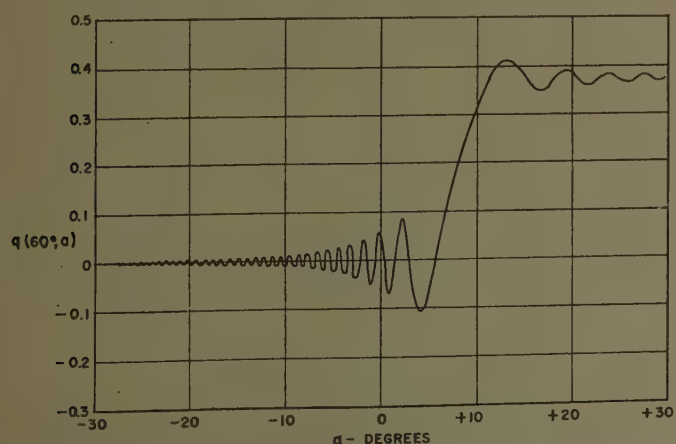


Fig. 15—Integral $q(\phi, a)$ for nonlinear phase function with $\phi = 60^\circ$.

A further comparison is made in the Appendix of the results for the 16-point Gaussian quadrature and the 58-point increment method in each of the ten 5.8° subintervals of the interval $(-29.0^\circ, 29.0^\circ)$. The values given in the Appendix for the two methods agree within better than one in the fourth decimal place for most subintervals. As seen above, this same order of agreement was obtained for the complete integral since the differences tend to cancel when the contributions from all subintervals are added together.

The results discussed in Section III-E for a linear phase function can be applied to the nonlinear phase functions in Fig. 9 to obtain an estimate of the largest subinterval size for which a 16-point Gaussian quadrature provides a satisfactory approximation. The degree of precision of the 16-point formula is $m=2n-1=31$. Assuming that the required degree of the power polynomial approximating $g(a, \phi)$ is about three times the number of roots in the subinterval, the number of cycles of $\Phi(a, \phi)$ in the subinterval must be limited to about five. This limit can be exceeded somewhat for those subintervals which make a very small contribution to the complete integral since a given percentage error in their contributions is a smaller percentage of the total. The results of Section III-E clearly indicate, however, that the error becomes large rapidly as this limit is exceeded.

The results presented in this Section satisfy the above condition fairly well except for $n_T=96$ which produced considerable error. The subinterval $(-29.0^\circ, -23.2^\circ)$ for $n_T=160$ has about 5.1, 6.2, and 6.8 cycles of $\Phi(a, \phi)$ for $\phi=0^\circ, 30^\circ$, and 60° , respectively. The subintervals $(-29.0^\circ, -19\frac{1}{3}^\circ)$, $(-29.0^\circ, -22\frac{5}{8}^\circ)$, and $(-29.0^\circ, -24\frac{1}{8}^\circ)$ corresponding to $n_T=96, 144$, and 192 have about 11.7, 7.9, and 5.9 cycles of $\Phi(a, 60^\circ)$, respectively.

V. DOUBLE NUMERICAL INTEGRATION

Double integration is performed numerically by applying one of the methods of single numerical integration discussed in Section II to a set of N ordinates which are the result of single numerical integrations with respect to one variable for N appropriate values of the second variable. Consider the double antenna pattern integral

$$E(u, v) = \int_c^d \int_{a(y)}^{b(y)} g(u, v, x, y) dx dy \quad (26)$$

where $g(u, v, x, y)$ equals either

$$g_1(u, v, x, y) = f(x, y) \cos \Phi(u, v, x, y)$$

or

$$g_2(u, v, x, y) = f(x, y) \sin \Phi(u, v, x, y).$$

The amplitude function, $f(x, y)$, and the phase function $\Phi(u, v, x, y)$, correspond to either the field distribution over an aperture or the current distribution over a reflector surface. In either case, the amplitude function may represent one component of a vector distribution and include any dot products that usually appear before

the integral sign. The parameters u and v are either the actual or normalized antenna pattern angles.

Eq. (26) can be written as

$$E(u, v) = \int_c^d I(u, v, y) dy \quad (27)$$

where

$$I(u, v, y) = \int_{a(y)}^{b(y)} g(u, v, x, y) dx. \quad (28)$$

Now let (28) be integrated numerically N times for N appropriate values of y , the j th numerical integration being

$$\begin{aligned} I(u, v, y_j) &= \int_{a(y_j)}^{b(y_j)} g(u, v, x, y_j) dx \\ &\doteq \sum_{i=1}^n H_{ij} g(u, v, x_i, y_j). \end{aligned} \quad (29)$$

Numerical integration of (27) then becomes

$$E(u, v) \doteq \sum_{j=1}^N W_j I(u, v, y_j) \quad (30)$$

where the W_j are the appropriate weight coefficients for the N -point numerical integration formula being used.

Eq. (29) and (30) can be combined to form the double summation

$$E(u, v) \doteq \sum_{j=1}^N \sum_{i=1}^n W_j H_{ij} g(u, v, x_i, y_j). \quad (31)$$

When a and b are functions of y as in (28), it is generally simpler to program the computer to perform the two operations given by (29) and (30) in sequence. When a and b are constants, on the other hand, the N sets of H_{ij} are the same for all j and (31) may be used conveniently.

In order to determine the values of n and N required for double numerical integration, the phase function can be calculated over the x - y plane for several combinations of (u, v) in the range to be used. Examination of the number of cycles of $\Phi(u, v, x, y)$ in the interval (a, b) for different values of y will determine the required value of n . While the number of cycles of $\Phi(u, v, x, y)$ in the interval (c, d) for different values of x will give some indication of the required value of N , this is more accurately determined from the number of oscillations of $I(u, v, y)$. If $\Phi(u, v, x, y)$ has regions of stationary phase when plotted against x for constant values of y (similar to the phase functions in Fig. 9), the number of oscillations of $I(u, v, y)$ in the interval (c, d) is about the same as the number of cycles through which the phase of the stationary phase points varies over the interval.

VI. CONCLUSIONS

The general principles of numerical integration have been discussed and four particular methods studied in detail. The increment method, Simpson's rule, and Gaussian quadratures resulted in about the same cost

per ordinate for an antenna pattern integral calculated on a digital computer. The higher degree of precision of the Gaussian quadrature results in a lower total cost per integral for that method, however, since fewer ordinates are required for a given accuracy. While the Gaussian quadrature has an error that becomes large rapidly for pattern angles greater than the range which yields good accuracy, it has the advantage that the error over most of this range is smaller than that obtained by the other two methods for a given maximum error. Proper application of Gaussian quadratures to antenna pattern integrals results, therefore, in greater accuracy and lower cost for a given maximum error than is obtained with the increment method or Simpson's rule.

When the antenna pattern integral has a linear phase function, Filon's method can be used and results in an error that is less than any desired amount over an essentially unlimited range of pattern angle. The maximum error occurs near the pattern angle u_2 for which there is one half cycle of the integrand per subinterval and is determined by the number of ordinates used. If the range of pattern angle is limited to about u_2 , the Gaussian quadrature will provide greater accuracy over most of the range at lower cost for the same number of ordinates, since Filon's method requires additional calculation time per integral to evaluate the $\alpha(u)$, $\beta(u)$, and $\gamma(u)$ coefficients. For a range of pattern angle appreciably greater than u_2 , however, Filon's method has the greater economy since the error remains below the maximum with no increase in the number of ordinates required.

The procedure for applying the methods of single numerical integration to double antenna pattern integrals has been presented. Different methods can be used for the two integrations with the number of ordinates required for each, n and N , being determined by the functions involved and the desired accuracy. Since the total number of ordinates required is n times N , it is important to use the most suitable method for each integration to achieve the greatest economy.

APPENDIX

Comparison of results for 16-point Gaussian quadrature and 58-point increment method in each 5.8° subinterval of the interval $(-29.0^\circ, 29.0^\circ)$.

| Subinterval | $p(0^\circ)$ | $q(0^\circ)$ |
|----------------|--------------|--------------|
| -29.0 to -23.2 | 0.000 562 | 0.001 546 |
| | 0.000 565 | 0.001 564 |
| -23.2 to -17.4 | -0.001 833 | -0.005 815 |
| | -0.001 859 | -0.005 875 |
| -17.4 to -11.6 | 0.009 540 | 0.003 078 |
| | 0.009 608 | 0.003 108 |
| -11.6 to -5.8 | -0.025 144 | 0.036 121 |
| | -0.025 252 | 0.036 264 |
| -5.8 to 0.0 | 0.249 268 | -0.255 671 |
| | 0.249 301 | -0.255 797 |
| 0.0 to 5.8 | 0.249 268 | -0.255 671 |
| | 0.249 301 | -0.255 797 |

| 5.8 to 11.6 | -0.025 144 | 0.036 121 |
|----------------|---------------|---------------|
| | -0.025 252 | 0.036 264 |
| 11.6 to 17.4 | 0.009 540 | 0.003 078 |
| | 0.009 608 | 0.003 108 |
| 17.4 to 23.2 | -0.001 833 | -0.005 815 |
| | -0.001 859 | -0.005 875 |
| 23.2 to 29.0 | 0.000 562 | 0.001 546 |
| | 0.000 565 | 0.001 564 |
| Subinterval | $p(30^\circ)$ | $q(30^\circ)$ |
| -29.0 to -23.2 | 0.000 486 | 0.003 884 |
| | 0.000 496 | 0.003 967 |
| -23.2 to -17.4 | -0.003 578 | -0.000 646 |
| | -0.003 645 | -0.000 662 |
| -17.4 to -11.6 | 0.004 524 | 0.005 684 |
| | 0.004 605 | 0.005 761 |
| -11.6 to -5.8 | 0.001 178 | 0.015 617 |
| | 0.001 195 | 0.015 721 |
| -5.8 to 0.0 | -0.096 633 | 0.012 830 |
| | -0.096 815 | 0.012 696 |
| 0.0 to 5.8 | 0.579 140 | 0.269 806 |
| | 0.579 285 | 0.269 770 |
| 5.8 to 11.6 | 0.043 663 | -0.110 027 |
| | 0.043 657 | -0.110 109 |
| 11.6 to 17.4 | -0.008 673 | 0.024 400 |
| | -0.008 692 | 0.024 458 |
| 17.4 to 23.2 | 0.001 845 | -0.005 054 |
| | 0.001 844 | -0.005 072 |
| 23.2 to 29.0 | 0.005 295 | 0.003 409 |
| | 0.005 322 | 0.003 428 |
| Subinterval | $p(60^\circ)$ | $q(60^\circ)$ |
| -29.0 to -23.2 | 0.000 412 | 0.000 822 |
| | 0.000 421 | 0.000 813 |
| -23.2 to -17.4 | 0.001 113 | -0.000 219 |
| | 0.001 142 | -0.000 254 |
| -17.4 to -11.6 | -0.003 638 | -0.006 416 |
| | -0.003 738 | -0.006 567 |
| -11.6 to -5.8 | 0.009 462 | -0.009 672 |
| | 0.009 598 | -0.009 784 |
| -5.8 to 0.0 | -0.011 961 | 0.075 809 |
| | -0.012 080 | 0.076 333 |
| 0.0 to 5.8 | 0.146 198 | -0.070 814 |
| | 0.146 227 | -0.071 093 |
| 5.8 to 11.6 | -0.068 686 | 0.389 808 |
| | -0.068 723 | 0.389 816 |
| 11.6 to 17.4 | 0.053 442 | -0.023 223 |
| | 0.053 442 | -0.023 237 |
| 17.4 to 23.2 | 0.001 685 | 0.021 343 |
| | 0.001 684 | 0.021 356 |
| 23.2 to 29.0 | 0.013 345 | -0.009 239 |
| | 0.013 359 | -0.009 251 |

VII. ACKNOWLEDGMENT

The author wishes to express his appreciation to A. E. Glow, Jr., for assistance during this study, to Miss C. J. MacDonald for programming the computer, and to E. F. Magnusson for suggesting the use of Gaussian quadratures.

The Numerical Solution of Antenna and Scattering Problems*

GEORGE SINCLAIR†

Summary—It is pointed out that numerical solutions of electromagnetic boundary-value problems deserve to receive greater emphasis, particularly those arising in practical applications. Numerical methods of solution can also be of great help in theoretical studies where suitable analytical solutions are difficult to obtain. By using a certain fundamental integral equation for the electromagnetic field, it is shown that many problems can be reduced to the solution of standard types of integral equations for which numerical techniques of solution already exist.

INTRODUCTION

THE purpose of the present study was to investigate the usefulness of numerical methods of solution of antenna and scattering problems. It was prompted by the need for more general methods of solving the boundary-value problems which are encountered in the design of practical electromagnetic systems. Also, there are reasons for believing that numerical solutions have much to contribute to theoretical studies of antenna and scattering problems when analytical solutions are difficult to obtain. Numerical solutions appear to be of definite interest in investigations of scattering by bodies in the so-called "resonance region."

In attempting solutions of boundary-value problems in electromagnetic theory, the emphasis is practically always placed on the desirability of obtaining analytical solutions. Numerical solutions of such problems are generally completely ignored, and this is most unfortunate as there are many situations where a numerical solution could provide as much (or even more) information than an analytical solution. Analytical solutions may be aesthetically more pleasing than numerical solutions, but this is of little interest to the engineer concerned with the design of a specific electromagnetic system.

Even when analytical solutions can be obtained for electromagnetic problems, it is usually found that a great amount of calculation is required to obtain numerical values, as, for example, when wave functions in a given coordinate system have to be computed. In many cases, direct numerical solution of a suitable differential or integral equation can yield the same accuracy of result with little, if any, increase in the amount of computing required.

It may be assumed that most persons concerned with the solution of electromagnetic problems now have access to high-speed electronic computers. It appears that

few of them are aware of the potentialities of a high-speed computer as an aid in the solution of boundary-value problems in electromagnetic theory. This situation is due, in part, to the lack of suitable numerical techniques for solving such problems.

Of the possible methods which could be used to obtain numerical solutions to antenna and scattering problems, those based on integral equations would appear to be most useful. For one thing, the boundary conditions can be inserted into the integral equation before attempting the solution, so that if a solution is obtained, it automatically satisfies the boundary conditions. For a specific problem, various integral equations can be derived which may differ in their suitability for numerical solution, but for many applications, an integral equation basically due to H. A. Lorentz^{1,2} offers a number of advantages, especially that of flexibility.

THE FUNDAMENTAL INTEGRAL EQUATION

The following integral equation is considered to be fundamental for analytical as well as numerical solutions of electromagnetic boundary-value problems. Its usefulness in the solution of a variety of problems was particularly pointed out by Albert and Synge.² Consider a region V bounded by a surface S in which a field exists, defined by the vectors \mathbf{E} and \mathbf{H} , and let \mathbf{E}_1 , \mathbf{H}_1 be another arbitrary field, of the same frequency in the same region. Then

$$\int_S (\mathbf{E}_1 \times \mathbf{H} - \mathbf{E} \times \mathbf{H}_1) \cdot \mathbf{n} dS \\ = \int_V (\mathbf{E} \cdot \mathbf{J}_1 - \mathbf{E}_1 \cdot \mathbf{J} - \mathbf{H} \cdot \mathbf{M}_1 + \mathbf{H}_1 \cdot \mathbf{M}) dv \quad (1)$$

(see Appendix).

The field \mathbf{E}_1 , \mathbf{H}_1 , which will be called the auxiliary field, is chosen to simplify the calculations. It is often taken to be a Green's function, but in many problems it will be more convenient to take a different choice, especially when the appropriate Green's function is not readily available.

The fundamental integral equation can be applied in a number of ways in order to solve a given problem, as has been indicated by Albert and Synge.² Typical meth-

* This work was performed at the University of Illinois, Urbana, Ill., while the author was on sabbatical leave from the University of Toronto.

† University of Toronto, Toronto, Ontario, Can.

¹ P. Frank and R. von Mises, "Differential- und integralgleichungen der mechanik und physik," vol. 2, Vieweg, Brunswick; 1935. Reprinted Rosenberg, New York, N. Y., p. 576; 1943.

² G. E. Albert and J. L. Synge, "The general problem of antenna radiation and the fundamental integral equation, with application to an antenna of revolution—part I," *Quart. Appl. Math.*, vol. 6, pp. 117–131; July, 1948.

ods of applying it will be illustrated by considering first a transmission problem and then a receiving (diffraction) problem. In each case, it will be shown that there results a linear integral equation of the first kind,

$$\int_a^b K(x, y)f(x)dx = g(y) \quad (2)$$

where $f(x)$ is the unknown function. A number of methods for the numerical solution of this type of integral equation can be found in the literature.³

A TRANSMITTING PROBLEM

By way of illustration, consider a two-dimensional problem, requiring the calculation of the pattern of an infinitely long narrow-slot antenna in a uniform infinite cylinder of arbitrary cross-sectional shape, when the slot is excited by a voltage applied across the gap formed by the slot. Suppose that the curve bounding the cross section is a closed curve (see Fig. 1).

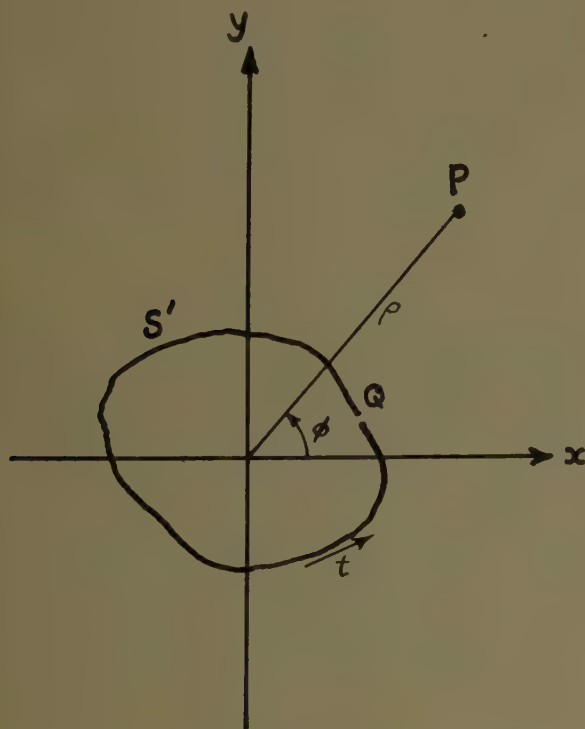


Fig. 1.—Coordinate system.

The choice of the region V is made as follows: let V be the region bounded by the surface of the cylinders, the surface at infinity, and contained between two transverse planes, whose separation may be chosen as unity without loss of generality. Assuming that both fields satisfy the radiation condition at infinity and have no variation in the axial or z -direction, then the surface integrals over the surface at infinity and over the transverse planes all vanish, leaving a surface integral

only over the surface S' of the cylinder. Noting that \mathbf{J} and \mathbf{M} are zero in V and that the tangential component of \mathbf{E} vanishes on S' , except at the gap where it equals V_0/δ (V_0 =voltage applied to gap of infinitesimal peripheral width δ), then

$$\int_{S'} E_{1t} H_z dS' - V_0 H_{1z}(Q) = - \int_V (\mathbf{E} \cdot \mathbf{J}_1 - \mathbf{H} \cdot \mathbf{M}_1) dv \quad (3)$$

where

E_{1t} = the component of E_1 tangential to the surface S' (see Fig. 1),

H_z = the axial component of \mathbf{H} on S' ,

$H_{1z}(Q)$ = the value of H_{1z} at the location of the slot.

In this equation, the unknowns are H_z in the surface integral and \mathbf{E} and \mathbf{H} in the volume integral. In order to solve the equation, it will be necessary to reduce the unknown quantities to one. Hence, assume that \mathbf{J}_1 and \mathbf{M}_1 are zero in V , so that the only sources for the auxiliary field lie *outside* the region V ; *i.e.*, lie *inside* the cylindrical surface S' . Then the integral equation becomes

$$\oint E_{1t} H_z dS = V_0 H_{1z}(Q) \quad (4)$$

where

s = distance along the periphery of the cross section from an arbitrary origin.

If this equation can be solved for a suitable choice of auxiliary field, then H_z on the surface of the cylinder (*i.e.*, the surface current produced by the voltage applied to the slot) is obtained.

The auxiliary field may be chosen arbitrarily, subject only to the restrictions already imposed, namely:

- 1) it must satisfy Maxwell's equations;
- 2) it must be of the same frequency as the field being determined;
- 3) it must satisfy the radiation condition at infinity;
- 4) it must be independent of z .

Eq. (4) is true for any auxiliary field satisfying these conditions. However, when this equation is being used to determine the unknown field H_z , it is important to consider the completeness of the representation chosen for H_z . The auxiliary field must be chosen in such a way as to allow for the determination of all the unknown parameters in the representation for H_z . This can be illustrated by considering the case of a circular cylinder. The appropriate choice for H_z on S' is then a Fourier series in the angular variable ϕ .

$$H_z = \sum_{m=-\infty}^{+\infty} c_m e^{im\phi} \quad (5)$$

³ Z. Kopal, "Numerical Analysis," Chapman & Hall, Ltd., London, Eng., Ch. 8; 1955.

A possible choice for the auxiliary field would be the cylindrical wave function

$$H_{1z} = H_p^{(2)}(k\rho)e^{ip\phi} \quad (6)$$

where p is a fixed integer. The integral equation would then determine only the single coefficient c_{-p} . The complete expression for H_z is obtained only when the parameter p is allowed to take on all possible integral values.

An alternative choice for the auxiliary field is the field of a uniform infinite line source of magnetic current located at a fixed distance d from the origin, in an arbitrary direction ϕ_1 . The usual addition theorem for cylindrical wave functions shows that this field has the required arbitrariness to yield all the coefficients in (5). For this case, the integral equation can be written in the form

$$\int_0^{2\pi} E_{1\phi}(\phi_1, \phi) H_z(\phi) a d\phi = V_0 H_{1z}(\phi_1) \Big|_{\phi=\phi'} \quad (7)$$

where ϕ' is the angular location of the slot on the cylinder. It is evident that this equation is of the same form as (2). It can be shown in a similar way that (4), for an arbitrary cylindrical surface, is also of the same form as (2) when the proper choice of auxiliary field is made.

Having determined the surface current, the radiated field produced at a given point P can be found by another application of the integral equation. Consider an auxiliary field produced by a line magnetic current K_0 volts located at the point P where the field from the slot is to be found. Then the integral equation becomes

$$\oint E_{1z}' H_z d\mathcal{S} = V_0 H_{1z}'(Q) + K_0 H_z(P) \quad (8)$$

where E' , H' is the field due to the line source at P . In this equation the only unknown is $H_z(P)$, so the field at P can be found by mechanical quadrature.⁴ If P lies in the far field of the slot, the pattern of the slot antenna is obtained by choosing P successively in various directions at fixed distance from the slot.

A RECEIVING OR DIFFRACTION PROBLEM

An alternative procedure for calculating the pattern of the slot antenna is to consider the receiving case. It is well known⁵ that the pattern of the slot antenna can be calculated if the field produced at the surface of the cylinder when a plane wave is incident from various directions can be found; *i.e.*, if the diffraction problem can be solved.

Let the cylindrical surface without any gap be designated S'' , and suppose the cylinder is illuminated by the field of an infinite magnetic line source K_0' volts which is located at a distant point P (it is convenient to use a line source instead of a plane wave in order to en-

sure vanishing of the surface integral over the surface at infinity). Assume that the sources for the auxiliary field lie outside V . Then, in this case, the integral equation becomes

$$\int_{S''} E_{1z} H_z dS'' = -K_0' H_{1z}(P). \quad (9)$$

In this equation the only unknown quantity is H_z on the surface S'' . If the equation can be solved for P located at a fixed distance in various directions, the pattern can be found by evaluating H_z at the point Q where the slot is to be located. This procedure is equivalent to calculating the pattern of the slot antenna by determining the current produced in a short circuit across the terminals of the antenna when a (plane) wave is incident from various directions. Eq. (9) is again of the same type as (2).

THREE-DIMENSIONAL PROBLEMS

The above techniques may be applied to problems in three dimensions. The amount of computing to be done may be greater than in two-dimensional problems, but not necessarily so. If the system has rotational symmetry, then the integral equation can usually be reduced to one of the same type as (2). In more general cases, it may be possible to solve the integral equation by analytical methods for one of the surface variables.

CONCLUSION

The fundamental integral equation, (1), has already proved to be very useful in analytical studies of electromagnetic boundary-value problems. It appears to be equally useful for numerical solutions of a wide variety of problems, particularly those which are difficult to treat by known analytical methods. It offers flexibility in the solution of a given problem, in that the auxiliary field can be chosen to simplify the mathematics as much as possible. For the engineer, it has the merit that the one integral equation, properly applied, provides a mathematical approach to a wide variety of problems.

APPENDIX I

THE LORENTZ RECIPROCITY THEOREM

Let \mathbf{E} , \mathbf{H} be the field associated with a distribution of electric and magnetic currents of densities \mathbf{J} and \mathbf{M} , and let \mathbf{E}_1 , \mathbf{H}_1 be another field of the same frequency, associated with a distribution of currents of densities \mathbf{J}_1 and \mathbf{M}_1 . Then Maxwell's equations can be written (time variation is assumed in the form $e^{j\omega t}$) as

$$\text{curl } \mathbf{H} = \mathbf{J} + j\omega\epsilon\mathbf{E}, \quad (10)$$

$$\text{curl } \mathbf{E} = -\mathbf{M} - j\omega\mu\mathbf{H}, \quad (11)$$

$$\text{curl } \mathbf{H}_1 = \mathbf{J}_1 + j\omega\epsilon\mathbf{E}_1, \quad (12)$$

$$\text{curl } \mathbf{E}_1 = -\mathbf{M}_1 - j\omega\mu\mathbf{H}_1. \quad (13)$$

⁴ *Ibid.*, Ch. 7.

⁵ G. Sinclair, "The patterns of slotted-cylinder antennas," *PROC. IRE*, vol. 36, pp. 1487-92; December, 1948.

Then by forming scalar products, the equations can be combined to yield

$$\begin{aligned} (E_1 \cdot \text{curl } H - H \cdot \text{curl } E_1) - (E \cdot \text{curl } H_1 - H_1 \cdot \text{curl } E) \\ = E_1 \cdot J + H \cdot M_1 - E \cdot J_1 - H_1 \cdot M. \quad (14) \end{aligned}$$

Hence

$$\begin{aligned} \text{div } (E_1 \times H - E \times H_1) \\ = E \cdot J_1 - E_1 \cdot J - H \cdot M_1 + H_1 \cdot M. \quad (15) \end{aligned}$$

Now consider a region V in which both fields exist, and let S be the surface enclosing the region. Then, integrating the above equation throughout the region V , and using the divergence theorem, there results

$$\begin{aligned} \int_S (E_1 \times H - E \times H_1) \cdot n dS \\ = \int_V (E \cdot J_1 - E \cdot J - H \cdot M_1 + H_1 \cdot M) dv \quad (16) \end{aligned}$$

where n is a unit normal directed out of the volume V . This is the expression of the reciprocity theorem to be used.

ACKNOWLEDGMENT

It is a pleasure to acknowledge the fact that the research was made possible by a Guggenheim grant. Discussions with Dr. W. V. Tilston and with personnel at the University of Illinois Antenna Laboratory were very helpful.

The Finite Conical Antenna*

S. ADACHI†, R. G. KOUYOUMJIAN‡, AND R. G. VAN SICKLE‡

Summary—This paper describes the results of a theoretical and experimental study of the finite conical antenna. The antenna consists of a perfectly-conducting cone of finite length excited at its tip by a linear element. The radiation patterns and radiation resistance have been calculated by taking the current distribution on the linear element to be sinusoidal and approximating the current distribution on the finite cone with that of an infinite cone. Calculated and measured radiation patterns are found to be in good agreement. The dependence of beamwidth, beam tilt, and side-lobe level on the antenna dimensions is discussed. A simple interpretation of the patterns in terms of the superposition of a wave radiated directly from the linear element (in the presence of the infinite cone) and a wave diffracted from the edge of the cone is pointed out. Measured values of the input VSWR (voltage standing-wave ratio) are presented and discussed.

INTRODUCTION

THIS paper describes the results of a theoretical and experimental study of the finite conical antenna. The antenna consists of a perfectly-conducting cone of finite length, excited at its tip by a linear element. The radiation patterns of the wide-angle conical antennas have been calculated by using the current distribution of an infinite cone as an approximation to the distribution on a finite cone. A similar approximate method has been applied quite successfully to the problem of the radiation from annular slots on a finite circular disc by Meixner.¹ The radiation field due to the finite cone ($0 \leq r \leq R_0$) is calculated as the difference between the radiation field due to the infinite cone ($0 \leq r \leq \infty$) and that due to the complementary cone ($R_0 \leq r \leq \infty$) with the infinite cone current. The radiation field due to the infinite cone antenna, and the current distribution on the complementary cone are obtained by making use of a special Green's function.² A sinusoidal current distribution is assumed along the linear element at the tip of the cone. The length of the cone is assumed to be large compared to a wavelength and the length of the antenna at the tip of the cone. The approximate expression for the far-zone electric field of the wide-angle conical antenna is obtained in rather simple form for computation. Calculated and measured radiation patterns are found to be in good agreement. The dependence of beam tilt, beamwidth and side-lobe

level on the antenna dimensions is shown, and the radiation resistance of the infinite, wide-angle conical antenna is obtained. This provides a reasonably good approximation to the radiation resistance of the finite conical antenna when the cone is large compared with a wavelength. The input VSWR of the finite conical antenna for different exciting elements at the tip of the cone is measured.

THEORETICAL FORMULATION

The configuration of the finite conical antenna is shown in Fig. 1, where it is seen that l is the length of

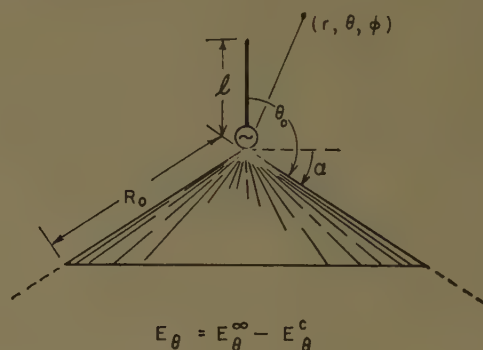


Fig. 1—Finite conical antenna.

thin linear element, R_0 is the slant length of the finite cone, and the conical surface is defined either by the angle θ_0 measured from the axis of the cone or the angle α measured from the horizontal plane. The antenna terminals are located at the junction of the linear element and cone tip. A sinusoidal current distribution is assumed to flow on the thin linear element, and as mentioned in the Introduction, the current distribution on the finite conical surface is approximated by that of the infinite cone. This may be expressed in an alternative way which is more convenient for carrying out the solution. Let the far-zone electric fields in the cases of an infinite cone, and a finite cone be denoted as E_θ^∞ and E_θ , respectively, and the far-zone electric field due to the infinite-cone current distribution flowing on the complementary base cone extending from R_0 to ∞ be denoted as E_θ^c . Then the far-zone electric field under consideration is approximately

$$E_\theta \simeq E_\theta^\infty - E_\theta^c. \quad (1)$$

The fields of an infinite conical antenna excited at its tip by a linear element are expressed by using a special Green's function which has been given by Felsen.² In addition to satisfying the usual requirements imposed on a Green's function, this Green's function vanishes on

* The research reported in this paper was sponsored in part by the U. S. Army Signal Res. and Dev. Lab., Ft. Monmouth, N. J.

† Antenna Lab., Dept. Elec. Engrg., The Ohio State University, Columbus, Ohio. Currently on leave of absence from Tohoku University, Sendai, Japan.

‡ Antenna Lab., Dept. Elec. Engrg., The Ohio State University, Columbus, Ohio.

¹ J. Meixner, "The radiation pattern and induced current in a circular antenna with an annular slit," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-4, pp. 408-411; July, 1956.

² L. B. Felsen, "Back-scattering from wide angle and narrow angle cones," *J. Appl. Phys.*, vol. 26, pp. 138-151; February, 1955.

the surface of the cone; furthermore, its representation in series form has the desirable property of rapid convergence when either the source point or the field point are close to the tip of the cone. The far-zone electric field has the form:³

$$E_{\theta}^{\infty}(r, \theta) = \frac{\eta I_0}{4} \frac{e^{-i\beta r}}{r} \sum_q e^{i(\pi/2)q} (2q+1) S_q(\beta l) \times \frac{P_q(-\cos \theta_0)}{\sin q\pi \frac{\partial}{\partial q} P_q(\cos \theta_0)} \frac{\partial}{\partial \theta} P_q(\cos \theta). \quad (2)$$

The magnetic field on the surface of the cone for $r > l$ is:

$$H_{\phi}^{\infty}(r', \theta_0) = -\frac{jI_0}{4r'} \sum_q (2q+1) S_q(\beta l) \times \frac{P_q(-\cos \theta_0) \frac{\partial}{\partial \theta} P_q(\cos \theta_0)}{\sin q\pi \frac{\partial}{\partial q} P_q(\cos \theta_0)} \hat{h}_q^{(2)}(\beta r') \quad (3)$$

where I_0 is the sinusoidal current amplitude on the linear element, η is the free space intrinsic impedance, and

$$S_q(\beta l) = \frac{1}{\beta} \int_0^l \frac{\sin \beta(l-r')}{r'^2} j_q(\beta r') dr' \quad (4)$$

$$j_q(x) = \sqrt{\frac{\pi x}{2}} J_{q+1/2}(x) \quad (5)$$

$$q: P_q(\cos \theta_0) = 0. \quad (6)$$

The far-zone electric field E_{θ}^c is given in terms of the magnetic field H_{ϕ}^{∞} on the complementary cone ($R_0 \leq r \leq \infty$) by

$$E_{\theta}^c = \frac{j\eta\beta}{4\pi} \frac{e^{-i\beta r}}{r} \int H_{\phi}^{\infty}(r', \theta_0) \bar{i}_2 \cdot (\bar{i}_1 \times \bar{r}_0 \times \bar{i}_1) e^{i\beta r' \bar{r}_0 \cdot \bar{i}_1} dS \quad (7)$$

where \bar{i}_1 and \bar{i}_2 are unit vectors in the directions of r and θ , respectively, and \bar{i}_0 is a unit vector directed from the origin toward the source point on the complementary cone surface. The integration is carried out over the complementary cone surface extending to infinity. We note that the above expression is approximate in the sense that it does not explicitly take into account the change of the unit vector directed from sources on the complementary screen to the field point. Although it is difficult to establish the validity of this approximation in the general case, it is readily done for the case where $\theta_0 = \pi/2$. Substituting (3) in (7) and performing the vector operation about the conical geometrical structure shown in Fig. 2, the E_{θ}^c is finally expressed in the form:

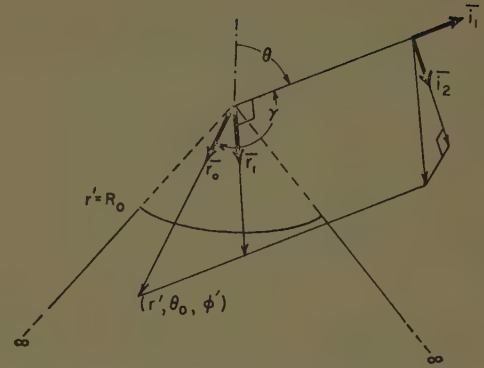


Fig. 2—Configuration and coordinate system for the infinite, complementary cone.

$$E_{\theta}^c(\theta) = \sin \theta_0 \frac{I_0 \eta}{16\pi} \frac{e^{-i\beta r}}{r} \sum_q (2q+1) S_q(\beta l) \times \frac{P_q(-\cos \theta_0) \frac{\partial}{\partial \theta} P_q(\cos \theta_0)}{\sin q\pi \frac{\partial}{\partial q} P_q(\cos \theta_0)} \times \int_{\beta R_0}^{\infty} d(\beta r') \hat{h}_q^{(2)}(\beta r') e^{i\beta r' \cos \theta_0 \cos \theta} \times \int_0^{2\pi} (\sin \theta_0 \cos \theta \cos \phi' - \cos \theta_0 \sin \theta) e^{i\beta r' \sin \theta_0 \sin \theta \cos \phi'} d\phi'. \quad (8)$$

Using the expansion:

$$e^{it \cos \phi'} = J_0(t) - 2J_2(t) \cos 2\phi' + 2J_4(t) \cos 4\phi' - \dots + j2\{J_1(t) \cos \phi' - J_3(t) \cos 3\phi' + \dots\} \quad (9)$$

the integration with respect to ϕ' in (8) can be easily carried out, i.e.,

$$\int_0^{2\pi} (\sin \theta_0 \cos \theta \cos \phi' - \cos \theta_0 \sin \theta) e^{i\beta r' \sin \theta_0 \sin \theta \cos \phi'} d\phi' = -2\pi [\cos \theta_0 \sin \theta J_0(\beta r' \sin \theta_0 \sin \theta) - j \sin \theta_0 \cos \theta J_1(\beta r' \sin \theta_0 \sin \theta)]. \quad (10)$$

The length of the cone is assumed to be large compared with the wavelength and the length of the linear element, and hence, the modified spherical Hankel functions in (3) can be expressed by their asymptotic forms:

$$\hat{h}_q^{(2)}(\beta r') \approx (j)^{q+1} e^{-i\beta r'}. \quad (11)$$

Substitution of (10) and (11) in (8) leads to

$$E_{\theta}^c(\theta) = E_{\theta}^{\infty}(\theta_0) \Theta(\theta), \quad (12)$$

where $E_{\theta}^{\infty}(\theta_0)$ is the far-zone electric field on the surface of the infinite cone, which is obtained by introducing $\theta = \theta_0$ into (2), and the function $\Theta(\theta)$ is expressed by the following integrations:

³ S. Adachi, "A Theoretical Analysis of Infinite Conical Antennas—Part I," Antenna Lab., The Ohio State Univ. Res. Foundation, U. S. Army Signal Res. and Dev. Lab., Fort Monmouth, N. J., Rept. 662-20, prepared under Contract DA 36-039 sc70174; February 15, 1959.

$$\Theta(\theta) = -\frac{\sin \theta_0}{2} \left[\sin \theta_0 \cos \theta \int_{\beta R_0}^{\infty} J_1(x \sin \theta_0 \sin \theta) \times e^{-jx(1-\cos \theta_0 \cos \theta)} dx \right. \\ \left. + j \cos \theta_0 \sin \theta \int_{\beta R_0}^{\infty} J_0(x \sin \theta_0 \sin \theta) \times e^{-jx(1-\cos \theta_0 \cos \theta)} dx \right]. \quad (13)$$

The factor $\Theta(\theta)$ is only a function of the configuration of the finite cone, and it is independent of the sinusoidal current element at the tip of the cone.

The condition

$$\beta R_0 \sin \theta_0 \sin \theta \gg 1 \quad (14)$$

is true in the case of a long ($\beta R_0 \gg 1$), wide-angle cone except for aspects θ close to the axis of the cone (θ close to 0 or π). Under this condition, Bessel functions of zeroth and first order in (13) can be expressed by the following well-known asymptotic forms:

$$J_0(\beta r \sin \theta_0 \sin \theta) \approx \sqrt{\frac{2}{\pi \beta r \sin \theta_0 \sin \theta}} \cos \left\{ \beta r \sin \theta_0 \sin \theta - \frac{\pi}{4} \right\} \\ J_1(\beta r \sin \theta_0 \sin \theta) \approx \sqrt{\frac{2}{\pi \beta r \sin \theta_0 \sin \theta}} \cos \left\{ \beta r \sin \theta_0 \sin \theta - \frac{3\pi}{4} \right\}. \quad (15)$$

Introducing (15) into (13) and carrying out the integration, the approximate expression for $\Theta(\theta)$ in terms of Fresnel's integrals is obtained:

$$\Theta(\theta) \approx -\frac{j}{2} \sqrt{\frac{\sin \theta_0}{\sin \theta}} [\cos \frac{1}{2}(\theta + \theta_0) \{1 - C(\xi \beta R_0) - S(\xi \beta R_0) - j(C(\xi \beta R_0) - S(\xi \beta R_0))\} \\ \pm \cos \frac{1}{2}(\theta - \theta_0) \{C(\zeta \beta R_0) - S(\zeta \beta R_0) - j(1 - C(\zeta \beta R_0) - S(\zeta \beta R_0))\}], \quad (16)$$

where the signs (\pm) correspond to the regions ($\theta \leq \theta_0$), respectively, and

$$\xi = 1 - \cos(\theta + \theta_0), \\ \zeta = 1 - \cos(\theta - \theta_0), \quad (17)$$

and $C(x)$ and $S(x)$ are Fresnel's integrals, i.e.,

$$C(x) - jS(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \frac{\cos t}{\sqrt{t}} dt \\ - j \frac{1}{\sqrt{2\pi}} \int_0^x \frac{\sin t}{\sqrt{t}} dt. \quad (18)$$

Thus, the far-zone electric field of the tip-excited finite cone is given by

$$E_\theta(r) = E_\theta^\infty(r) - E_\theta^\infty(r_0)\Theta(\theta). \quad (19)$$

Next, consider the $E_\theta^\infty(r, \theta)$ expressed by (2). For the

wide angle cone, the following approximations can be made:

$$q \approx \frac{(n + \frac{1}{2})\pi - \theta_0}{2\theta_0} \quad (20)$$

where n is an odd integer, and consequently

$$\frac{P_q(-\cos \theta_0)}{\sin q\pi \frac{\partial}{\partial q} P_q(\cos \theta_0)} \approx -\frac{1}{\theta_0}. \quad (21)$$

Substitution of (21) in (2) yields

$$E_\theta^\infty(r, \theta) \approx -\frac{\eta I_0}{4\theta_0} \frac{e^{-j\beta r}}{r} \sum_q e^{j(\pi/2)q} (2q+1) S_q(\beta l) \frac{\partial P_q(\cos \theta)}{\partial \theta}. \quad (22)$$

It can be shown by the equivalence between an electric dipole and a small annular slot on an infinite plane that in the special case of an infinitesimally small dipole on a circular disc ($\theta_0 = \pi/2$), (19) is identical to the expression for the field from a small annular slot on a circular disk, as discussed by Meixner.¹

As mentioned before with regard to (14), the asymptotic expression of $\Theta(\theta)$ in (16) is invalid for high angle radiation. The two integrals in (13) must be carried out by numerical integration in the region of θ close to 0 and π . The integrations from βR_0 to ∞ are transformed into the integrations from 0 to $\beta R_0(1 - \cos \theta_0 \cos \theta)$, as given in the Appendix.

It is known^{4,5} that the radiation resistance of a linear antenna on a large circular disk is fairly well approximated by the value of the radiation resistance of the same linear antenna located on an infinite ground plane. The radiation resistance of an infinite, wide-angle cone excited by the linear antenna is obtained in the following form:

$$R \approx \frac{\eta}{4\theta_0 \sin^2 \beta l} \sum_q q(q+1)(2q+1) S_q(\beta l)^2. \quad (23)$$

DISCUSSION

The radiation patterns have been calculated for the finite cones: 1) $\theta_0 = 90^\circ$, 103.81° , 115.50° , and 2) $\beta R_0 = 15$, 30, 50. The function $S_q(\beta l)$ given by (4) is rather complicated for arbitrary values of βl .³ When $\beta l = (\pi/2)i$, $i = 1, 2, 3, \dots$, $S_q(\beta l)$ is given simply by

$$S_q\left(\frac{\pi}{2}i\right) = \frac{j_q\left(\frac{\pi}{2}i\right)}{q(q+1)}. \quad (24)$$

The calculations presented here will be made only for

⁴ J. E. Storer, "The impedance of an antenna over a large circular screen," *J. Appl. Phys.*, vol. 22, pp. 1058-1066; August, 1951.

⁵ A. S. Meier and W. P. Summers, "Measured Impedances of Vertical Antennas over Finite Ground Planes," Antenna Lab., The Ohio State University, Columbus, Rept. 233-3; 1946.

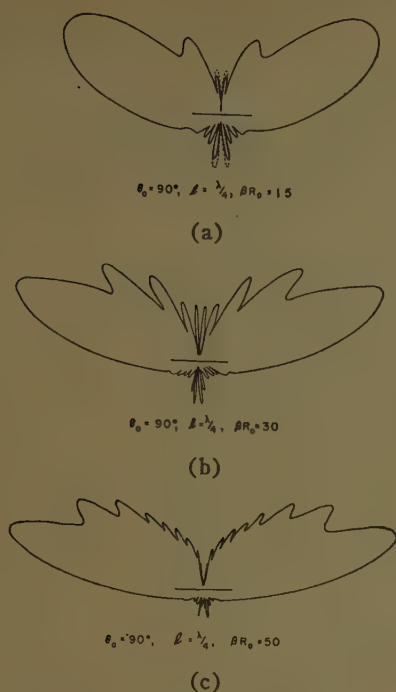


Fig. 3—Calculated patterns.

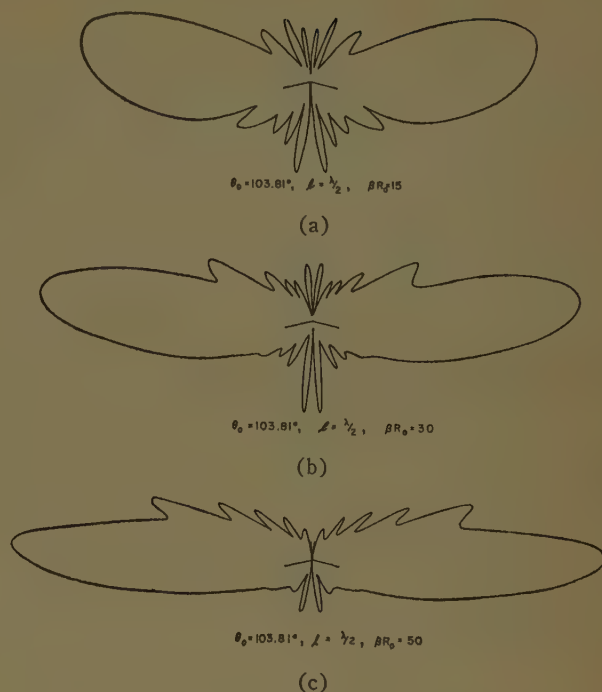


Fig. 5—Calculated patterns.

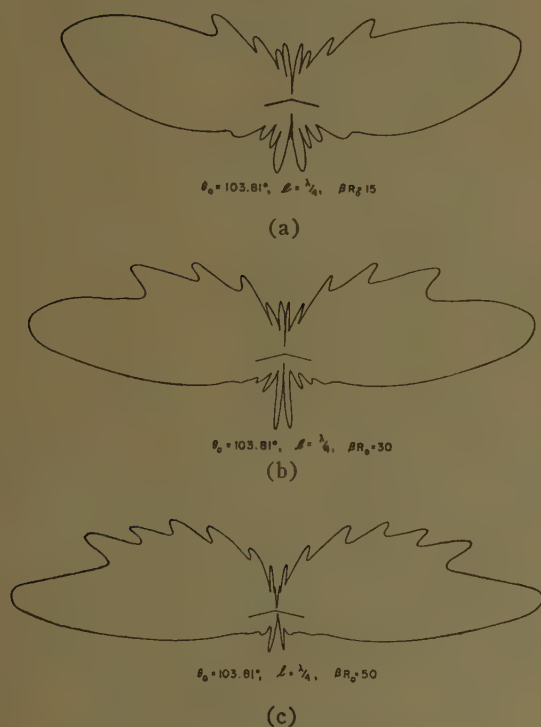


Fig. 4—Calculated patterns.

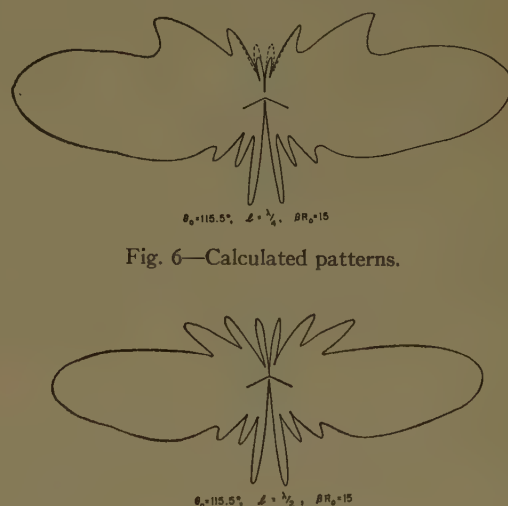


Fig. 6—Calculated patterns.

Fig. 7—Calculated pattern.

$l = \lambda/4$, and $\lambda/2$. The radiation pattern of an antenna over a circular ground screen ($\theta_0 = 90^\circ$) has been theoretically treated by use of spheroidal wave functions⁶ and by a variational method.⁷ The former solution is not appropriate for a large ground screen because of the

lengthy numerical computation, and the latter one is not available for the numerical computation of the entire pattern.

Calculated, polar, radiation patterns are shown in Figs. 3–7. The disk is considered here to be a special case of the finite cone ($\theta_0 = 90^\circ$). In each figure the orientation of the pattern with respect to the conical surface is indicated by a sketch at the center of the pattern. The characteristics of the antenna such as the direction of the main beam, the width of the main beam and the side-lobe level are readily obtained from these patterns. In Figs. 3–5, the slant length of the cone given by BR_0 increases in the division of the figures into (a), (b), and (c). As the slant length increases, it is seen that the direction of the main beam more closely approaches the conical surface and that the sidelobes on the upper part

⁶ A. Leitner and R. D. Spencer, "Effect of a circular ground-plane on antenna radiation," *J. Appl. Phys.*, vol. 21, pp. 1001–1006; October, 1950.

⁷ J. E. Storer, "The radiation pattern of an antenna over a circular ground screen," *J. Appl. Phys.*, vol. 23, pp. 588–593; May, 1952.

of the pattern increase in number and amplitude (relative to the main beam). In the case of the infinite cone the direction of maximum radiation would be along the conical surface and the ripple-like sidelobes would vanish. Comparing Figs. 4 and 5, it is seen that these sidelobes are less pronounced when the more directive $\lambda/2$ linear element replaces the $\lambda/4$ linear element. These effects may be explained by using the following simple, geometrical interpretation of the radiation from the antenna. Let the field of the antenna be regarded as the superposition of a wave radiated directly from the linear element in the presence of the infinite cone and a wave diffracted from the edge of the cone. This interpretation is based on (19), where $E_\theta^\infty(\theta)$ is the former component and $E_\theta^\infty(\theta_0)\Theta(\theta)$ is the latter component, with $\Theta(\theta)$ being naturally interpreted as a diffraction coefficient. Thus, for field points above the conical surface the two components interfere and give rise to the ripples on the pattern. Below the conical surface there is just one component present, and consequently the pattern is smooth. This behavior is typical of antennas where a significant portion of the radiation emanates from a wave guided over a finite distance. In this case, however, it should be noted that some of the lobe structure at aspects in the vicinity of $\Theta=0^\circ$ and 180° may be attributed to the interference of waves emanating from different segments of the circular edge.

The relationship between the direction of maximum radiation γ and the cone angle θ_0 , and the relationship between γ and the cone length are illustrated in Figs. 8 and 9, respectively. It is interesting to note that γ decreases almost linearly with increasing θ_0 and that γ becomes less sensitive to incremental variations in βR_0 and l as βR_0 increases. Fig. 8 provides a theoretical basis for the design of a finite conical antenna having a main beam in the direction $0 \sim \gamma \sim 30^\circ$.

The numerical integration of (25) and (26) for the high-angle radiation pattern ($\theta \approx 0, \pi$) has been carried out only for two cases because of the formidable computational difficulties. In the case of Fig. 6, only the part of the radiation pattern for $\theta \approx 0$ has been calculated. The results of the numerical integration are indicated by dotted lines, and it is seen that this more painstaking calculation does not lead to a great difference in the pattern. It can be shown that the difference is very small for the longer conical antennas.

In making the initial comparisons between calculated and measured patterns, conical antennas having the same βR_0 and l as those used in the calculations were not available. However, as has been implied earlier, small differences in these dimensions are not too important. Comparisons of Fig. 10(a) with Figs. 6 and 7 and Fig. 10(b) with Fig. 3(a), generally speaking, show good agreement between the calculated and measured patterns. In Fig. 11 the measured and calculated beam tilts γ are seen to be in very good agreement; conical antennas having the same dimensions as those used in the calculations were available for this comparison.

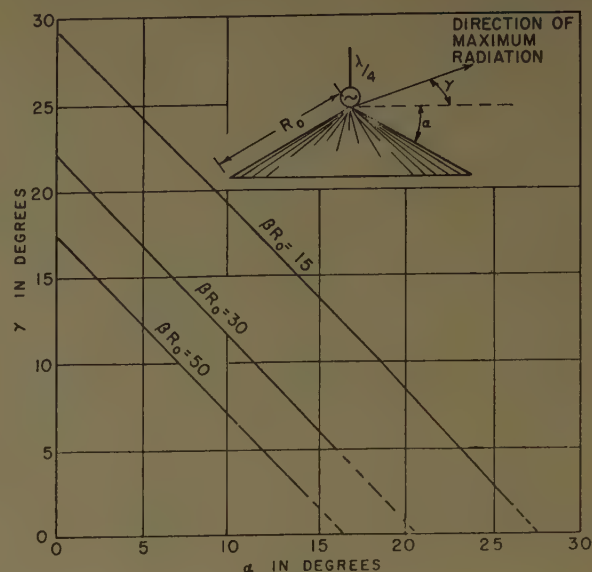


Fig. 8.

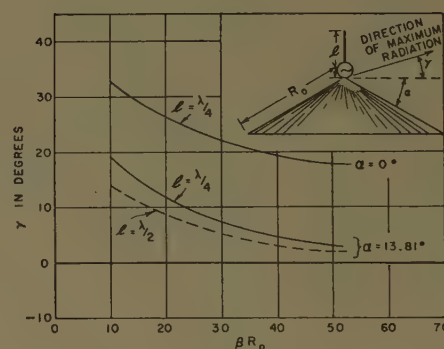


Fig. 9.

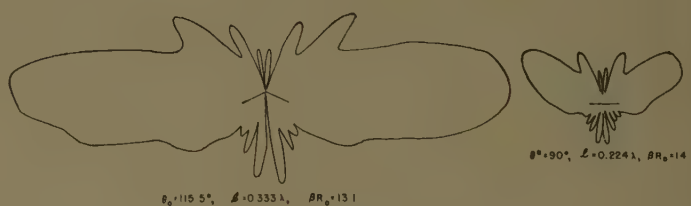


Fig. 10—Measured patterns.

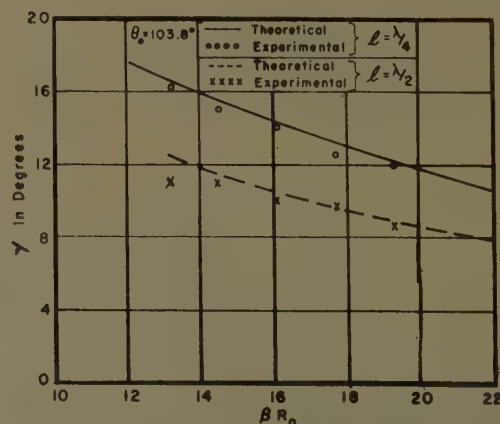


Fig. 11—Direction of maximum radiation.

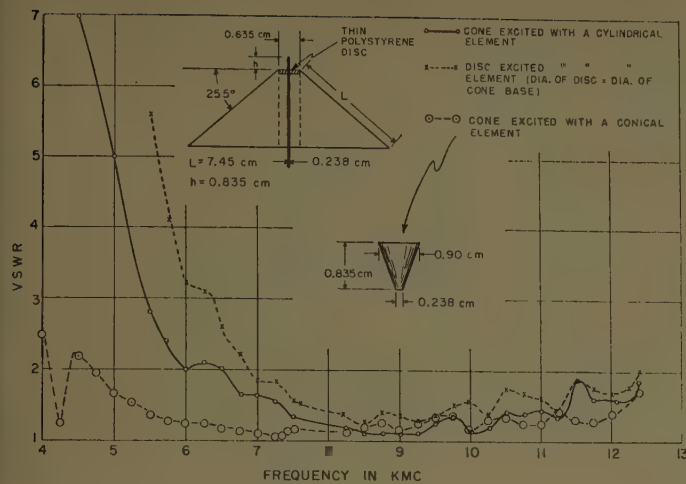
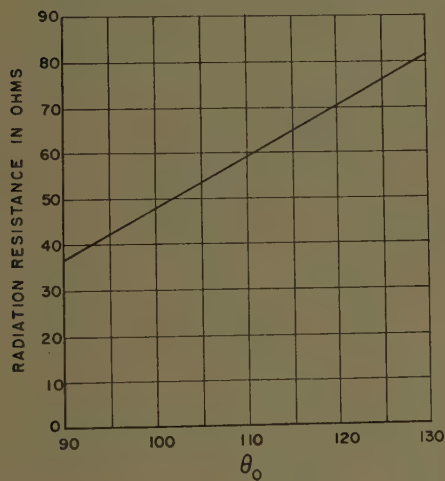


Fig. 12—Plot of VSWR vs frequency.

Fig. 13—Radiation resistance of an infinite, wide-angle cone (length of linear element = $\lambda/4$).

The input VSWR for the finite conical antenna with different exciting elements at its tip was measured, and is plotted in Fig. 12. The cone used in the experiment is made of a solid piece of aluminum with a hole drilled along its axis to accommodate the feeding coaxial cable. The dimensions of the cone are shown in Fig. 12; at 9 kmc the linear element is a quarter wavelength long and the slant length of the cone in terms of βR_0 is 14.1. The measurement for the disk is simply carried out by

inserting the cylindrical element through the base of the cone. The characteristic impedance of the coaxial cable is 57.5 ohms. The radiation resistance approximated by that of the infinite cone case ($\theta_0 = 115.5^\circ$; length of linear element = $\lambda/4$) is 65 ohms, as may be seen from Fig. 13. This result is consistent with the low VSWR observed in the vicinity of 9 kmc, and it also indicates that the reactive part of the input impedance is relatively low in this vicinity. If the VSWR curves for the disk and cone were plotted as a function of βR_0 , they would be nearly superimposed. The most interesting result, however, is that when the cylindrical exciting element is replaced by the conical element shown in the figure, there is no rapid increase in VSWR below 7 kmc, and the antenna has good broadband properties with respect to the input VSWR. It was noted in replacing the cylindrical exciting element with a conical exciting element that there is an inappreciable change in the pattern. Thus, to a first approximation the input VSWR of this antenna can be adjusted to an optimum without affecting the patterns.

APPENDIX: THE TRANSFORMATION OF TWO DEFINITE INTEGRALS

$$\int_{\beta R_0}^{\infty} J_0(x \sin \theta_0 \sin \theta) e^{-jx(1 - \cos \theta_0 \cos \theta)} dx$$

$$= \mp \frac{j}{\cos \theta - \cos \theta_0} - \frac{1}{1 - \cos \theta_0 \cos \theta}$$

$$\times \int_0^{\beta R_0(1 - \cos \theta_0 \cos \theta)} J_0\left(x \frac{\sin \theta_0 \sin \theta}{1 - \cos \theta_0 \cos \theta}\right) e^{-jx} dx \quad (25)$$

$$\int_{\beta R_0}^{\infty} J_1(x \sin \theta_0 \sin \theta) e^{-jx(1 - \cos \theta_0 \cos \theta)} dx$$

$$= \frac{1}{\sin \theta_0 \sin \theta} \left(1 \mp \frac{1 - \cos \theta_0 \cos \theta}{\cos \theta - \cos \theta_0} \right) - \frac{1}{1 - \cos \theta_0 \cos \theta}$$

$$\times \int_0^{\beta R_0(1 - \cos \theta_0 \cos \theta)} J_1\left(x \frac{\sin \theta_0 \sin \theta}{1 - \cos \theta_0 \cos \theta}\right) e^{-jx} dx, \quad (26)$$

where the signs (\mp) correspond to the regions ($\theta \leq \theta_0$), respectively.

Broadband Multislot Antenna

P. MARIE†

Summary—A multislot antenna was built, of which the VSWR is smaller than two in the band of 100 to 1000 mc.

The principle of this antenna is analogous to the one of pass band filters in which few modes of resonance are possible, appearing one after the other when frequency varies in the band. While one mode is decreasing and the following is increasing, the simultaneous transmitted waves by the two modes are in quadrature of phase.

To do this, it is necessary to have, for each central frequency of the different modes, the radiated electric field on the antenna in the opposite direction with respect to the direction of the electric field for the central frequencies of the two close modes, the direction of the electric field in the feeder being the same in every case.

THE purpose here is to show the possibility for multislot antennas of theoretically unlimited bandwidth but practically limited by mechanical considerations. This may be demonstrated as analogous to a lattice network.

Consider Fig. 1. This lattice network consists of two pairs of oscillating circuits, one pair turned on frequency f_1 and the second pair on frequency f_2 .

Let f_1 be smaller than f_2 .

Let R be the load and the characteristic impedance of the line by which the power is entering the filter.

Let Q be the quality factor of the tuned circuits at work.

$$Q = \frac{\sqrt{\frac{L_1}{C_1}}}{R} = \frac{\sqrt{\frac{L_2}{C_2}}}{R}.$$

Let the difference $(f_2 - f_1)$ be such that

$$\frac{f_2 - f_1}{f_2 + f_1} = \frac{1}{Q}.$$

Consider Fig. 2. It is seen by the dotted lines, as a function of the frequency, the transmitted power and the phases of the transmitted waves by each pair of tuned circuits work alone, the other pair of tuned circuits being out.

When the two pairs of tuned circuits are working together, the transmitted power and the phase are represented by full lines; the two pairs of tuned circuits are working independently at every frequency.

For all frequencies smaller than f_1 , the first pair of circuits is working practically alone. For all frequencies greater than f_2 , the second pair of circuits is working alone, and between these frequencies f_1 and f_2 the two

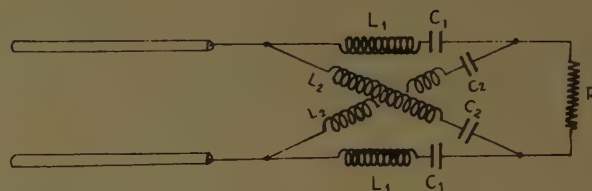


Fig. 1—Lattice network with two pairs of tuned circuits.

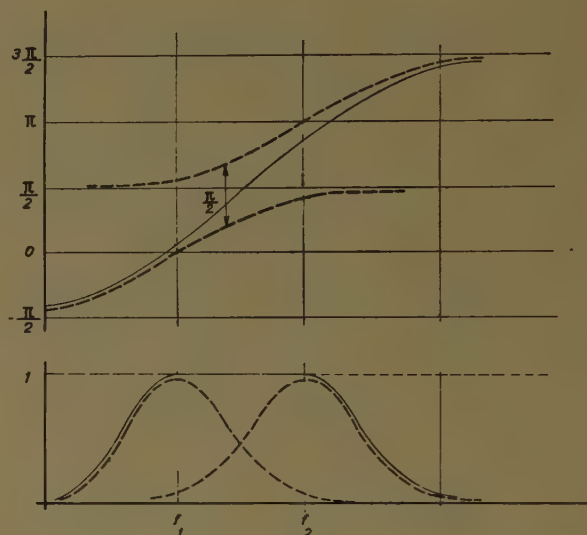


Fig. 2—Phase and power transmitted: ----, through each pair of tuned circuits working alone; —, through the two pairs of tuned circuits working together.

pairs of circuits are working independently, because in this region the difference of phase between the two transmitted waves is approximately constant and equal to $\pi/2$. Therefore, the two waves are independent of each other and we can add the electric fields and the powers.

The method may be extended to three or more tuned circuits. The number of tuned circuits would be limited only in practice, but not theoretically. If Q is given, the bandwidth of the filter is proportional to the number of pairs of tuned circuits. This method may be extended to multislot antennas.

Consider Fig. 3. Three slots are on a panel. Slots A and B are the same, tuned at f_2 frequency and excited in the same phase by two 100-ohm coaxial lines fed by a 50-ohm coaxial line represented by dotted lines. Slot C is tuned at f_1 frequency and excited by the motion of the electric charges on the edges of slots A and B nearest to slot C .

It is to be noted, with respect to the electric charges

† Centre National d'Etudes des Telecommunications, Issy-Les-Moulineaux, France.

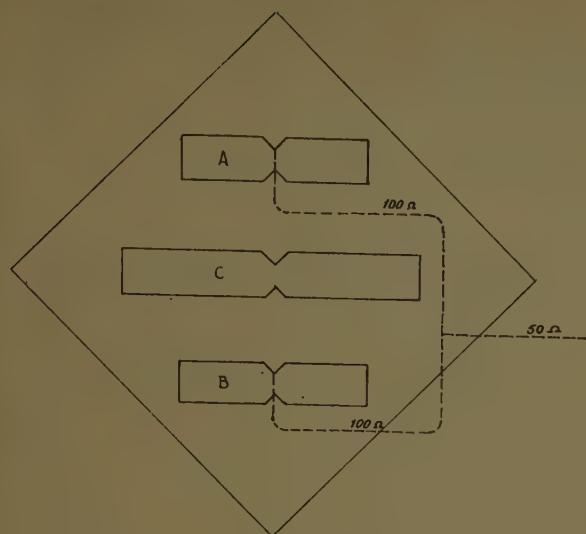


Fig. 3—A multislot antenna.

brought by the feeder, indicated + and - on Fig. 3, that the electric field in slot *C* would be just in the opposite direction to the field of slots *A* and *B*, if the phase would not be modified by the reactance of the slot.

So the excitation of slots *C* and *AB* are crossed just like the two pairs of oscillating circuits in Fig. 1.

When slots *C* and *AB* are radiating together, the radiated waves are in quadrature of phase, and electric field and power are added. Changing the frequency, we can go from f_1 to f_2 without discontinuity of the impedance at the transmitter. Note that the excitation of the panel working like a dipole is crossed with the excitation of slot *C* and thus the radiation of slot *C* is joined with the radiation of the panel and its dimensions give the lowest frequency radiated by our device.

In this way a multislot antenna was built with a VSWR of less than two, in a frequency range between 100 and 1000 mc.

Resonance and Supergain Effects in Small Ferromagnetically or Dielectrically Loaded Biconical Antennas

CHARLES POLK†

Summary—The input impedance and the radiation pattern of thin biconical antennas which contain materials of high dielectric permittivity ϵ and high magnetic permeability μ can be evaluated by methods due to Schelkunoff and Tai. Examination of numerical results shows that resonances of the input impedance are obtainable for antenna lengths much shorter than that of the ordinary half-wave dipole. It becomes also apparent, however, that the impedance bandwidth at these resonances is narrow and decreases with increasing ϵ and μ .

Analysis of the pertinent equations leads to the following conclusions concerning radiation patterns: 1) The pattern of an antenna which is small in terms of free-space wavelengths λ and is also small in terms of material wavelengths, $\lambda/\sqrt{\mu\epsilon}$, is essentially that of a small current element in air; 2) If the biconical antenna is small in terms of free-space wavelengths, but is large in terms of material wavelengths the radiation pattern is also that of a small current element, except in certain well defined and very narrow frequency bands where multi-lobe or narrow lobe patterns appear. A similar "supergain" effect has been noted by Knudsen [33] in his discussion of small loops which carry currents of varying phase.

INTRODUCTION

ANTENNAS containing dielectric or ferromagnetic materials have been investigated as early as 1946 [1-3]. With the advent of new materials having either very high dielectric constant (ferroelectrics) or having, simultaneously, fairly high dielectric constant and magnetic permeability (ferrites), interest in antennas of this type has been renewed [4-14] and, in particular, the question has been raised whether such antennas offer any advantages over other types in the HF and VHF range.

The two problems to be considered are the effects of the materials upon antenna input impedance and upon the radiation pattern. It is well known that the radiation pattern of a current element of infinitesimal length differs only slightly from the pattern of a half-wave dipole [15]; nevertheless, at frequencies where the physical size of the half-wave dipole does not become prohibitive, this larger antenna is decidedly more practical. The reason is that the input impedance of a small antenna consists essentially of a large reactive term. The radiation resistance is small compared with the input reactance [15-17] and the loss resistance of the matching circuit which must be inserted between the reactive antenna and the transmitter or receiver is likely to be larger than the radiation resistance; consequently a considerable part of the available power is wasted in the matching network.

The use of dielectric and ferromagnetic materials in small antennas seems to be desirable if it does not change the radiation pattern in such a way as to produce many lobes and if, at the same time, the input impedance is modified in such a manner as to reduce the reactive component without making the input resistance either impractically small or large for matching.

If an antenna without dielectric or ferromagnetic material is sufficiently small in terms of free-space wavelengths, only the small current element ("dipole") pattern will show up in the far field [18]. This is discussed later in the section called "Radiation Patterns" and in the Appendix. It will also be shown that a thin biconical antenna containing dielectric or ferromagnetic material, as indicated in Fig. 1, will also exhibit the pattern of a small dipole, if it is small in terms of free-space wavelengths and material wavelengths. If the antenna is small in terms of free-space wavelengths, but large in terms of material wavelengths, the dipole pattern will again be preserved, except at certain well defined frequencies, or lengths.

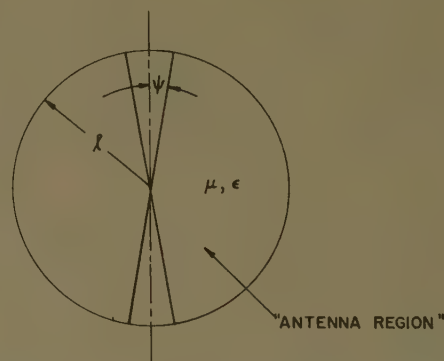


Fig. 1—Biconical antenna containing material with relative dielectric permittivity ϵ and relative magnetic permeability μ .

INPUT IMPEDANCE

To obtain numerical results, it is useful to consider the biconical antenna because this structure can be analyzed exactly and also because it reduces, in the limiting case, to the cylindrical antenna [20]. The input impedance of a thin biconical antenna embedded in dielectric material has been derived by Tai [21] and a slightly more general expression applicable to a biconical antenna embedded in a lossless material of arbitrary permeability and permittivity has been given by Schelkunoff [22]. Tai has also considered a biconical antenna embedded in a dissipative medium [23].

† Dept. of Elec. Engrg., University of Rhode Island, Kingston, R. I. Formerly with RCA Laboratories, Princeton, N. J.

The input impedance, Z_s , of a loss-free, symmetrical biconical antenna considered as a terminated uniform transmission line is

$$Z_s = \frac{K_1 Z_T + i K_1^2 \tan \beta_1 l}{K_1 + i Z_T \tan \beta_1 l} \quad (1)$$

where

K_1 = characteristic impedance of biconical transmission line,

$$\begin{aligned} K_1 &= 120 \sqrt{\frac{\mu}{\epsilon}} \log_e \left[\cot \frac{\psi}{2} \right] \\ &= \sqrt{\frac{\mu}{\epsilon}} K_0 = \frac{\eta_1}{\eta} K_0, \end{aligned} \quad (2)$$

μ, ϵ = relative permeability and permittivity,

ψ = apex semiangle of cone (Fig. 1),

K_0 = characteristic impedance of biconical line with air dielectric,

η, η_1 = wave impedances $\sqrt{\mu_0/\epsilon_0}$ and $\sqrt{\mu\mu_0/\epsilon\epsilon_0}$.

Z_T is the effective terminal impedance of the line when used as antenna; it is related to Schelkunoff's "inverse radiation impedance" $Z_a = K_1^2 Y_T$ (the impedance seen at the distance $\lambda/4$ from the end of the line) by

$$Z_T = \frac{K_1^2}{K_1^2 Y_T} = \frac{K_0^2}{K_0^2 Y_T} \quad (3)$$

Schelkunoff [22] and Tai [21] give expressions for $K_1^2 Y_T$ which are valid when ψ is small. By an obvious modification of their equations one obtains the expression for $K_0^2 Y_T$ for the symmetrical biconical antenna illustrated in Fig. 1 which contains material of relative permeability μ and relative dielectric permittivity ϵ .

$$\begin{aligned} K_0^2 Y_T = -i120 \sum_{k=1,3,5,\dots} \frac{2k+1}{k(k+1)} & \left[\frac{R_k'(\beta_0 l)}{R_k(\beta_1 l)} \right. \\ & \left. - \sqrt{\frac{\mu}{\epsilon}} \frac{S_k'(\beta_1 l)}{S_k(\beta_1 l)} \right]^{-1} \end{aligned} \quad (4)$$

where

β_0 = phase constant in free space = $\omega \sqrt{\mu_0 \epsilon_0}$

ω = radian frequency

$\beta_1 = \beta_0 \sqrt{\mu \epsilon}$

$$R_k(x) = \sqrt{\frac{\pi x}{2}} H_{k+1/2}^{(2)}(x), \quad R_k'(x) = \frac{d}{dx} R_k(x)$$

$$S_k(x) = \sqrt{\frac{\pi x}{2}} J_{k+1/2}(x), \quad S_k'(x) = \frac{d}{dx} S_k(x).$$

$H_{k+1/2}^{(2)}(x)$ and $J_{k+1/2}(x)$ are, respectively, the Hankel functions of the second kind and the Bessel functions with half-order index which usually occur in problems with spherical symmetry [24-26].

Numerical results for $K_0^2 Y_T$ were obtained to three significant figure accuracy on an IBM-650 computer [27, 28]. Results for three different values of β_1 ($\beta_1 = \beta_0$, $\beta_0 \sqrt{10}$, $\beta_0 10$) and lengths l corresponding to $0 \leq \beta_0 l \leq 1$ are shown on Fig. 2. Once values of $K_0^2 Y_T$ are available, one can easily compute actual values of input impedance of biconical antennas with any specified small apex semiangle by employing (1), (2) and (3). As an example, the input impedance of biconical antennas with apex angle $\psi = 2^\circ$ is plotted on Figs. 3 to 7 for $\mu = \epsilon = 1$; $\mu = 1$, $\epsilon = 10$; $\mu = 1$, $\epsilon = 28$; $\mu = \epsilon = 10$ and $\mu = 1$, $\epsilon = 100$.

RADIATION PATTERNS

The radiation field for a thin biconical antenna can be obtained from expressions given by Schelkunoff [22] as shown in the Appendix. For the antenna (see Fig. 1) with $\mu = \epsilon = 1$ it is

$$rE_\theta = ia \sum_{k=1,3,5,\dots} \frac{2k+1}{k(k+1)} R_k'(\beta_0 r) S_k(\beta_0 l) P_k'(\cos \theta), \quad (5)$$

where

a = constant defined in the Appendix,

r = distance from the antenna center to point where field is evaluated,

$P_k'(\cos \theta) = (d/d\theta) P_k(\cos \theta)$ = first associated Legendre function with index k .

The functions $P_k'(\cos \theta)$ with indexes $k = 1, 3, 5, 7$ are plotted on Figs. 8 to 10. $P_1'(\cos \theta) = \sin \theta$ is the single lobe pattern of the infinitesimal dipole, while the higher order functions exhibit many lobes. In the Appendix it is shown that $R_k'(\beta r)$ is a phase term which does not contribute to the magnitude of the terms in the series and that all except the first term are negligible when the antenna is small ($\beta_0 l \ll 1$).

When the "antenna region" in Fig. 1 is filled with dielectric or ferromagnetic material, (5) is replaced by

$$\begin{aligned} rE_\theta = a \sum_{k=1,3,5,\dots} \frac{2k+1}{k(k+1)} \frac{R_k'(\beta_0 r)}{R_k(\beta_1 l)} & \left[\frac{\eta}{\eta_1} \frac{R_k'(\beta_0 l)}{R_k(\beta_0 l)} \right. \\ & \left. - \frac{S_k'(\beta_1 l)}{S_k(\beta_1 l)} \right]^{-1} P_k'(\cos \theta). \end{aligned} \quad (6)$$

All symbols are defined below (2), (4) and (5). In the Appendix it is shown that for $\beta_0 l \ll 1$ and $\beta_1 l \ll 1$ all except the $k = 1$ term in the series are negligible. When $\beta_0 l \ll 1$, but $\beta_1 l \gg 1$, which is the case when materials with large μ and ϵ are used, it becomes apparent (see Appendix) that again all terms beyond the one with $k = 1$ are negligible, except for certain critical lengths or frequencies given by

$$\tan [\beta_0 l \sqrt{\mu \epsilon}] = \beta_0 l \sqrt{\mu \epsilon} = \beta_1 l \quad \beta_1 l \gg 1 \quad (7)$$

$$\sqrt{\frac{\epsilon}{\mu}} \frac{3}{\beta_0 l} = \frac{\beta_1 l \sin \beta_1 l + 6 \cos \beta_1 l}{\beta_1 l \cos \beta_1 l - 6 \sin \beta_1 l} \quad \beta_1 l \gg 1. \quad (7a)$$

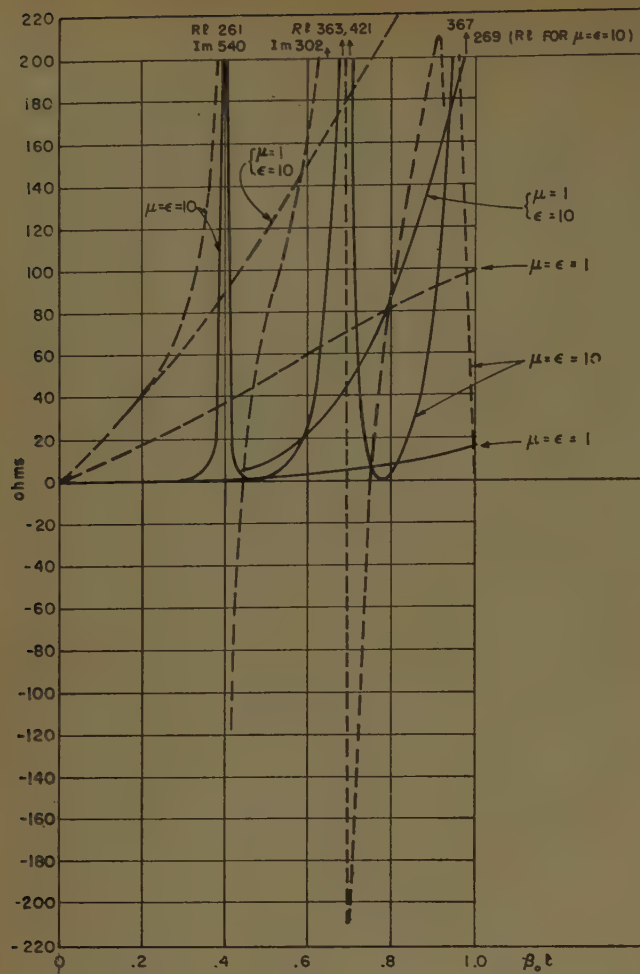


Fig. 2— $K_0^2 Y_T$ for 2° biconical antennas.

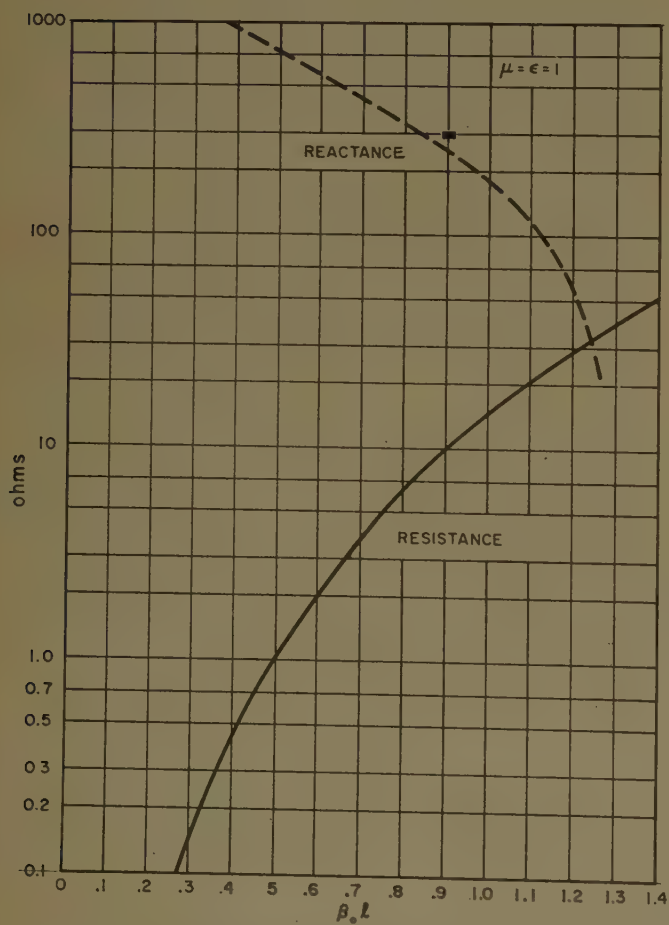


Fig. 3—Input impedance of 2° biconical antenna $\mu = \epsilon = 1$.

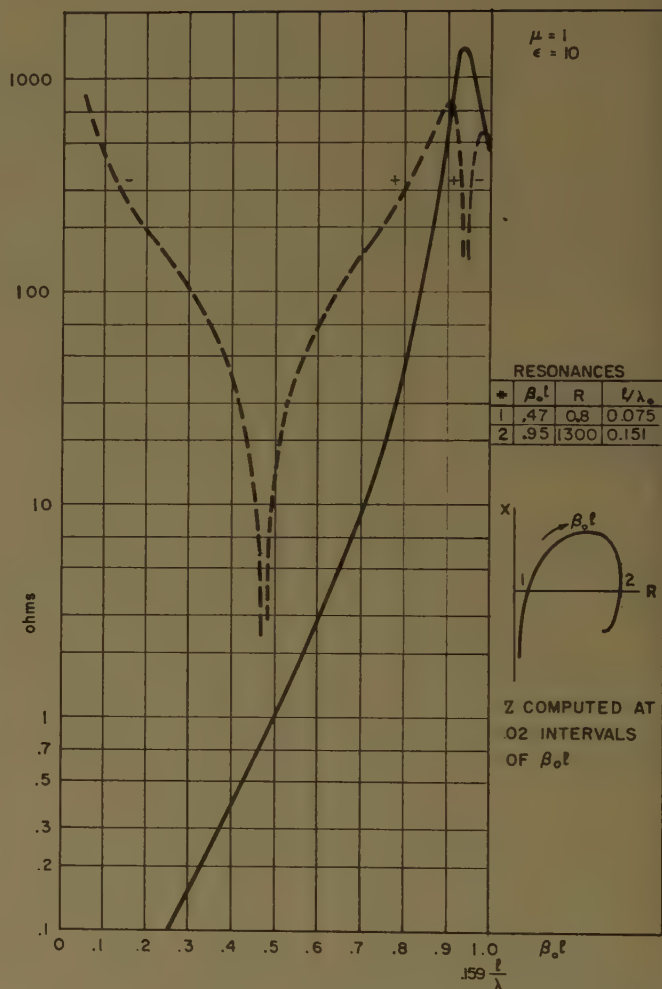


Fig. 4—Input impedance of 2° biconical antenna $\mu = 1$, $\epsilon = 10$.

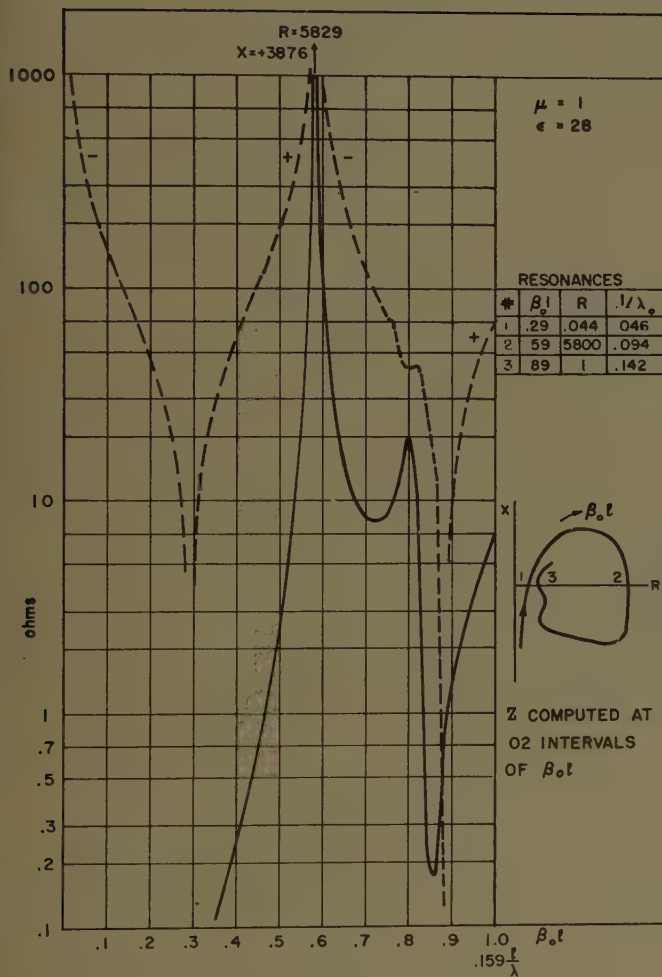


Fig. 5—Input impedance of 2° biconical antenna, $\mu=1$, $\epsilon=28$.

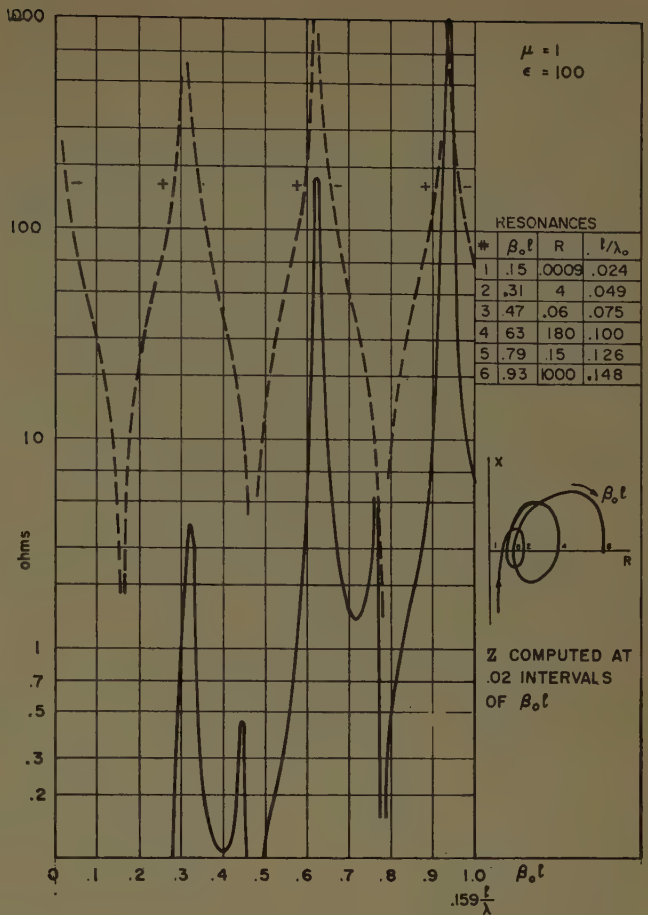


Fig. 6—Input impedance of 2° biconical antenna, $\mu=1$, $\epsilon=100$.

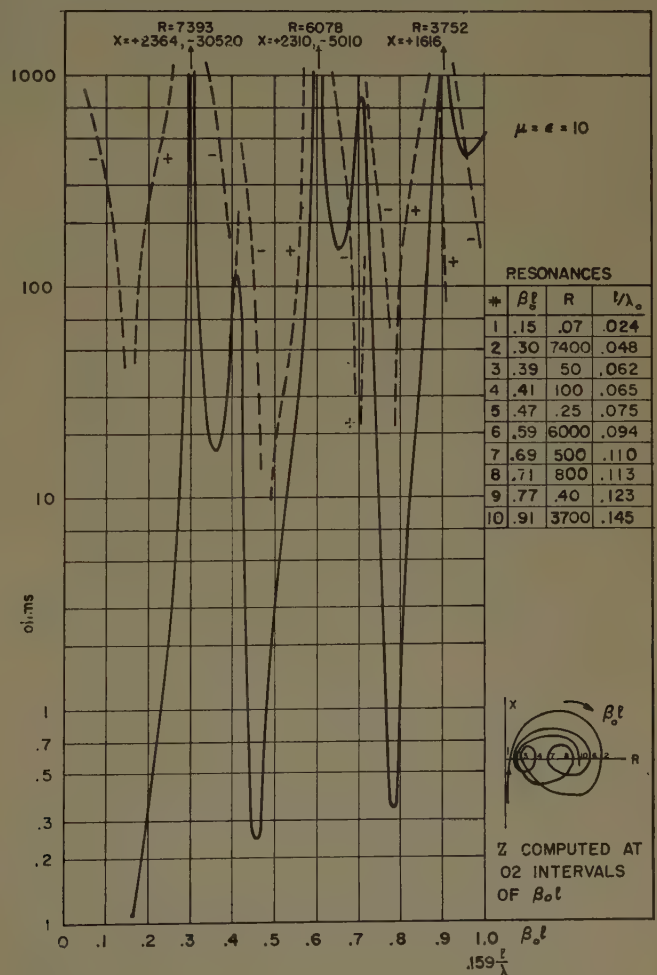


Fig. 7—Input impedance of 2° biconical antenna, $\mu=\epsilon=10$.

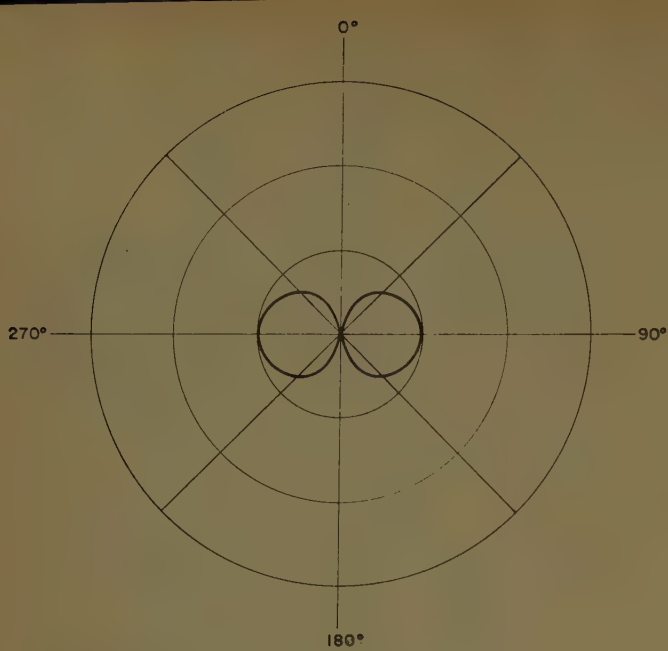


Fig. 8— $P_1'(\cos \theta)$.

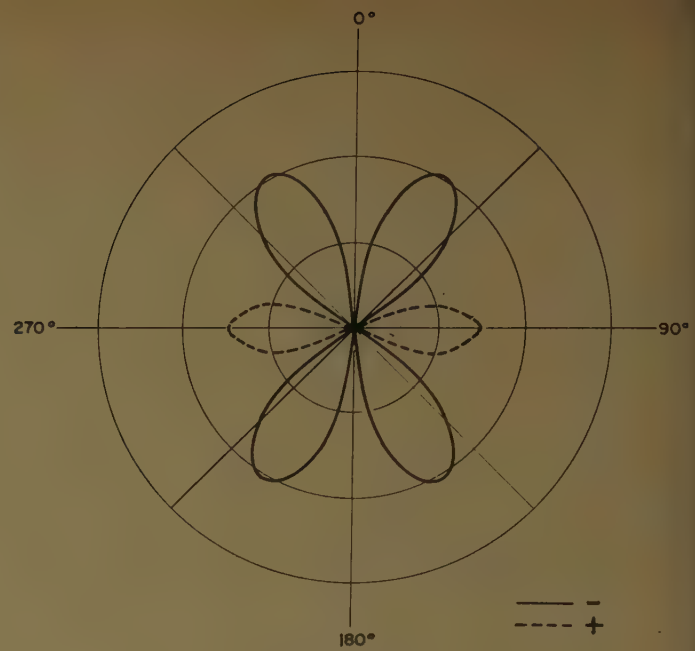


Fig. 9— $P_3'(\cos \theta)$.

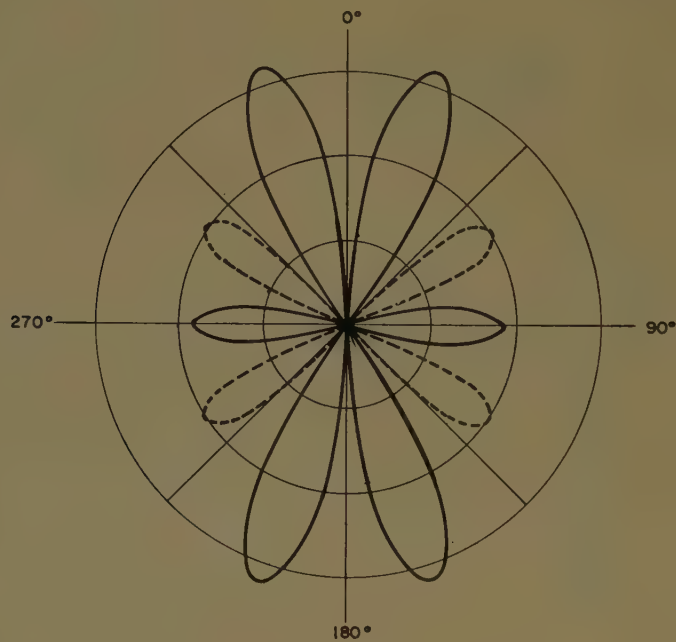


Fig. 10— $P_6'(\cos \theta)$.

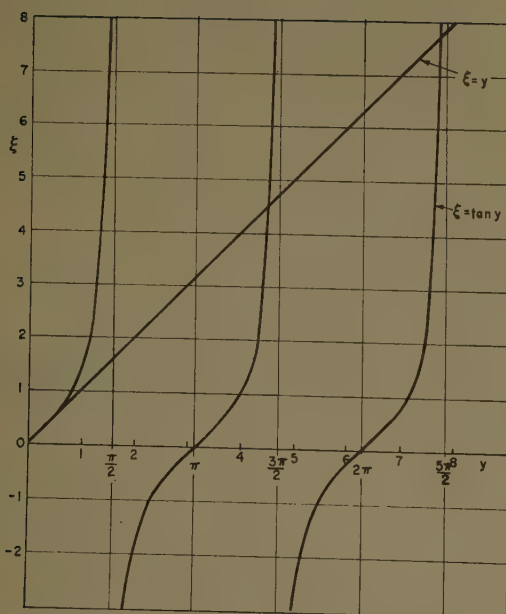


Fig. 11—Solutions of $\tan y = y$; intersections of $\xi = y$ and $\xi = \tan y$ are solutions.

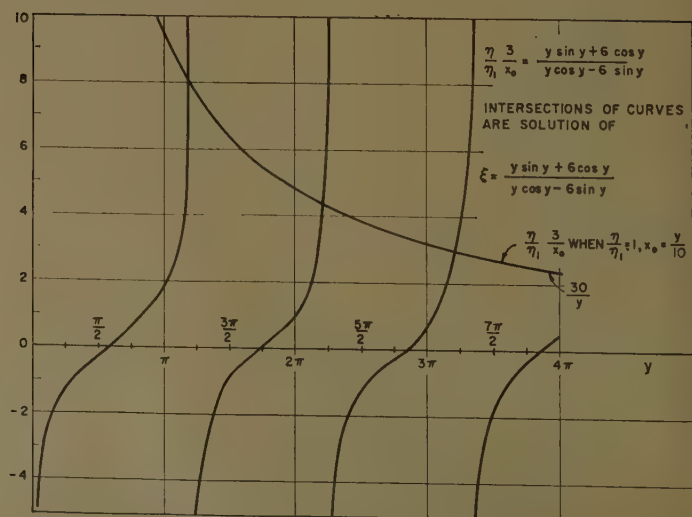


Fig. 12—Solution of (7a) or (30).

These equations can be solved graphically as shown on Figs. 11 and 12 for the special case $\mu\epsilon = 100$.

Comparison with Fig. 7 ($\eta/\eta_1=1$) and Fig. 6 ($\eta/\eta_1 \neq 10$) shows that these solutions always lie in the neighborhood of extreme values of the input impedance.

The result stating that all higher order terms in (6) are negligible in comparison with $P_1'(\cos \theta) = \sin \theta$ when the antenna is small in terms of free-space wavelengths ($\beta_0 l \ll 1$) and material wavelengths ($\beta_1 l \ll 1$) is hardly surprising. When the antenna is small in terms of free-space and material wavelengths only the fundamental TEM (transmission line) mode can be supported inside the antenna region, and, since the antenna is small, the current distribution must be nearly uniform. The radiation pattern should, therefore, be that of a short, uniform current element.

When the antenna is small in terms of free-space wavelengths ($\beta_0 l \ll 1$), but large in terms of material wavelengths ($\beta_1 l \gg 1$), the antenna region (see Fig. 1) corresponds to a long transmission line. The current distribution on it may therefore be highly nonuniform. The result expressed by (7) indicates that under these circumstances there exist critical frequencies (or lengths) when the small antenna ($\beta_0 l \ll 1$) will produce radiation patterns normally only associated with much larger antennas. At these frequencies it thus becomes a "supergain" antenna in the sense discussed by Chu [32], and one would expect extreme frequency dependence of both radiation patterns and impedance in the neighborhood of the "supergain" points. A similar result has been obtained by Knudsen [33] for a small circular radiator which carries a progressively phased current with the phase going through several complete reversals along the circumference.

CONCLUSION

The object of the computations was to obtain some quantitative idea of the effect which ferromagnetic and dielectric materials of finite extent would have when they surround a small dipole-type antenna. Examination of Figs. 3 to 7 shows the following.

1) Biconical antennas filled with dielectric or ferromagnetic material exhibit resonances at shorter lengths than the corresponding air-filled antennas. The first resonance, which in the air-filled antenna occurs in the neighborhood of $l = \lambda/4$ ($\beta_0 l = 1.57$), is shifted almost exactly according to the rule $(\beta_0 l)_{\text{resonance}} = 1.57/\sqrt{\mu\epsilon}$.

2) When $\epsilon = 10$, the radiation resistance (the resistive component of the input impedance) at first resonance is about the same as the resistance of the air-filled antenna of the same length. When $\mu = \epsilon = 10$, the resistance at first resonance is lower (approximately 0.1 ohm) than the corresponding value when $\epsilon = 10$, $\mu = 1$ (approximately 0.8 ohm), but higher than the value for $\epsilon = 100$, $\mu = 1$.

3) When μ or ϵ are sufficiently large, second and higher resonances occur at several lengths which are smaller than the length for first resonance ($\lambda/4$) of the

air-filled biconical antenna. While extremely large and extremely small values of input resistance (greater than 1000 ohms and less than 0.1 ohm) appear at some of these resonances, it is apparent from Figs. 5 to 7 that zero reactance points also exist where the input resistance lies between 0.1 and 1000 ohms.

4) The insertion of dielectric or ferromagnetic material into the biconical antenna structure causes increased frequency dependence of the antenna. The larger the product $\mu\epsilon$, the more rapid is the variation of the input impedance. This is apparent from Figs. 4 to 7, where it is illustrated qualitatively by the polar plots of input impedance which go through an increasing number of revolutions over the same frequency range (or range of $\beta_0 l$) as the values of μ and ϵ are increased. For example, considering first resonances, on the basis of the plotted data one obtains [29, 30] an effective unloaded antenna Q of 5.3 for the 2° air-filled biconical antenna (at $\beta_0 l = 1.57$); when $\epsilon = 10$ ($\beta_0 l = 0.47$) the Q lies between 95 and 240 and when $\mu = \epsilon = 10$ ($\beta_0 l = 0.15$) one obtains $2700 \leq Q \leq 5600$. A Q of 3000 at a frequency of 100 mc corresponds to a half-power width of 33 kc; ample for voice communication purposes, but much too narrow for television.

Resonance is obtained at shorter length clearly at the expense of decreased bandwidth. This is of course not objectionable for applications where narrow bandwidth is sufficient. One should, however, keep in mind that the largest obtainable value of μ or ϵ is not necessarily the best; rather the lowest values of μ , ϵ which will give the desired impedance at the desired length should be selected if bandwidth is at all a consideration. Also, in view of the pronounced dependence of the system upon μ and ϵ , stability of material properties (which may for example depend upon temperature) must be carefully evaluated when antennas of this type are to be constructed.

Finally, there is a pronounced effect of losses upon a high Q circuit. The introduction of a loss resistance equal in magnitude to the radiation resistance of the antenna at resonance (R_{res}) would reduce the antenna Q to one half of the corresponding value in the loss-free case. A loss resistance equal to R_{res} means, of course, that the power dissipated as heat would be equal to the radiated power. Such a loss resistance (equal to R_{res}) would, however, be only a very small fraction of the antenna reactance at a frequency only slightly removed from resonance in the numerical examples considered earlier; in other words, keeping in mind that

$$Q = \frac{\pi (\text{energy stored in circuit})}{\text{energy dissipated and radiated per half cycle}},$$

it is clear that in a high Q antenna where the value of the energy radiated per cycle is only a small fraction of the energy stored in the induction field, the ratio of lost power to radiated power will be large, if only a small fraction of the stored energy is converted into heat.

5) The radiation pattern of a small biconical antenna

($\beta_0 l \ll 1$) filled with dielectric or ferromagnetic material is identical with that of a small current element except at certain well defined points. There the antenna exhibits "supergain" characteristics, producing patterns associated normally only with much larger antennas; both patterns and impedance are highly frequency dependent in this case.

APPENDIX

EVALUATION OF RADIATION PATTERNS

The radiation field of a biconical antenna is given by Schelkunoff [22] as

$$rE_\theta = i \frac{\eta}{2\pi} \sum_{k=1,3,5,\dots} \frac{b_k}{k(k+1)} \frac{R'_k(\beta r)}{R_k(\beta l)} \frac{d}{d\theta} P_k(\cos \theta), \quad (8)$$

where $P_k(\cos \theta)$ is the Legendre function of index k and the derivative $(d/d\theta)P_k(\cos \theta)$ is identical [24] with the first associated Legendre function $P_k^m(\cos \theta)$.

R'_k and R_k are defined as above:

$$\begin{aligned} R_k(x) &= \sqrt{\frac{\pi x}{2}} H_{k+1/2}^{(2)}(x) \\ &= \sqrt{\frac{\pi x}{2}} [J_{k+1/2}(x) - iN_{k+1/2}(x)] \end{aligned} \quad (9)$$

$$R'_k(x) = \frac{d}{dx} R_k(x). \quad (10)$$

The coefficients b_k are given [22] for the "air-filled" antenna by

$$b_k = \frac{-V(l)}{K} (2k+1) \frac{\pi \beta l}{2} [J_{k+1/2}(\beta l)] [H_{k+1/2}^{(2)}(\beta l)] \quad (11)$$

where $V(l)$ is the voltage between the two cones of the antenna at the radial distance $r=l$, and K is the characteristic impedance of the antenna (2). Thus, from (8) and (11) the expression for the radiation field is

$$rE_\theta = -i \frac{\eta V(l)}{2\pi K} \sum_{k=1,3,5,\dots} \frac{2k+1}{k(k+1)} R'_k(\beta r) S_k(\beta l) P'_k(\cos \theta). \quad (12)$$

At points removed from the antenna $\beta r = y \gg 1$ one may therefore use the first term of the asymptotic expansion for $\sqrt{\pi y/2} H_{k+1/2}^{(2)}(y)$ [24]:

$$R_k(y) = \sqrt{\frac{\pi y}{2}} \sqrt{\frac{2}{\pi y}} \exp - i \left[y - \frac{k+1}{2} \pi \right]. \quad (13)$$

It is obvious, therefore, that

$$R'_k(\beta r) = -i \exp - i \left[\beta r - \frac{k+1}{2} \pi \right]$$

only contributes to the phase and not to the amplitude of the k th term in the summation of (12).

If the antenna is small in terms of free-space wavelengths such that $\beta l = x \ll 1$, the first term of the power series expansion for $J_{k+1/2}(x)$ may be used [24] in (12):

$$S_k(x) = \sqrt{\frac{\pi x}{2}} J_{k+1/2}(x) = \sqrt{\pi} \frac{x^{k+1}}{\Gamma(k+3/2)2^{k+1}}. \quad (14)$$

Clearly, only the first term in (12) contributes to the radiation field if the antenna is small. This, of course, is well known; $P'_1(\cos \theta) = \sin \theta$ is the radiation pattern of a linear current element.

The next question to be answered concerns the radiation pattern which is obtained if the "antenna region" of Fig. 1 is filled with dielectric ($\epsilon > 1$) or ferromagnetic ($\mu > 1$) material (if ferrite is used $\mu > 1$ and $\epsilon > 1$ are satisfied simultaneously).

The coefficients b_k in (8) must in this case be multiplied [22] by

$$1 + \frac{c_k}{b_k} = -i [S_k(\beta_1 l) R_k(\beta_1 l)]^{-1} \left[\frac{\eta}{\eta_1} \frac{R'_k(\beta_0 l)}{R_k(\beta_0 l)} - \frac{S'_k(\beta_1 l)}{S_k(\beta_1 l)} \right]^{-1}. \quad (15)$$

As before, β_0 is the phase constant in free space and $\beta_1 = \sqrt{\mu\epsilon} \beta_0$ is the phase constant in the antenna region. If the multiplication is performed, the expression for the radiation field becomes [from (8), (12) and (15)]

$$rE_\theta = - \frac{V(l)\eta}{2\pi K} \sum_{k=1,3,5,\dots} \frac{2k+1}{k(k+1)} \frac{R'_k(\beta_0 l)}{R_k(\beta_1 l)} \left[\frac{\eta}{\eta_1} \frac{R'_k(\beta_0 l)}{R_k(\beta_0 l)} - \frac{S'_k(\beta_1 l)}{S_k(\beta_1 l)} \right]^{-1} P'_k(\cos \theta). \quad (16)$$

If one notices the relation between Bessel functions [24] $J_\nu(x)N_\nu(x) - N'_\nu(x)J_\nu(x) = 2/\pi x$, it is easy to show that (16) reduces to (12) if $\beta_1 = \beta_0$ and $\eta = \eta_1$. As long as small antennas are considered such that $\beta_0 l \ll 1$, two cases are of interest:

1) $\beta_1 l \ll 1$. This condition places an upper bound upon μ and ϵ if $\beta_0 l$ is finite, since $\beta_1 l = \beta_0 l \sqrt{\mu\epsilon}$. The antenna is small not only in terms of the free-space wavelength λ_0 , but also in terms of the material wavelength $\lambda_0/\sqrt{\mu\epsilon}$.

2) $\beta_0 l \ll 1$ while $\beta_1 l \gg 1$. No upper bound is placed upon μ and ϵ . The antenna is small in terms of the free-space wavelength, but large in terms of the material wavelength.

If we want to decide whether the field pattern contains terms $P'_k(\cos \theta)$ with $k > 1$ which are significant, we are only interested in the behavior of the following expression in (16), since $|R'_k(\beta_0 r)| = 1$ (13):

$$f_k = \frac{2k+1}{k(k+1)R_k(\beta_1 l)} \left[\frac{\eta}{\eta_1} \frac{R'_k(\beta_0 l)}{R_k(\beta_0 l)} - \frac{S'_k(\beta_1 l)}{S_k(\beta_1 l)} \right]^{-1}. \quad (17)$$

The analysis of the two cases indicated above now requires different steps.

Case 1: ($\beta_0 l \ll 1, \beta_1 l \ll 1$)

Employing (14) and appropriate approximations for the Hankel [25] and Bessel [31] functions we obtain,

when $x_0 = \beta_0 l \ll 1$ and $x_1 = \beta_1 l \ll 1$,

$$f_k = -\frac{2k+1}{k(k+1)} \frac{\sqrt{\pi} x_1^k}{i\Gamma(k+\frac{1}{2})2^k} \cdot \left[\frac{\eta}{\eta_1} \frac{k}{x_0} + \frac{(k+1)}{x_1} \right]^{-1}. \quad (18)$$

The ratio of successive terms in (16) is then

$$\frac{|f_{k+2}|}{|f_k|} = \frac{(2k+5)(k+1)k}{(2k+1)^2(2k+3)(k+3)(k+2)} \cdot \frac{\eta k x_1 + \eta_1(k+1)x_0}{\eta(k+2)x_1 + \eta_1(k+3)x_0} x_1^2. \quad (19)$$

Obviously, when $x_1 \leq 1$, $x_0 \ll 1$,

$$\left| \frac{f_{k+2}}{f_k} \right| < 1. \quad (20)$$

For example, when

$$\eta = \eta_1, \quad x_0 = 0.1, \quad x_1 = 0.4:$$

$$|f_3/f_1| = 1.56(10^{-3}).$$

Case 2: ($\beta_0 l \ll 1$, $\beta_1 l \gg 1$)

Again, since the magnitude of $R_k'(\beta_0 r)$ in (16) is one, only the factor f_k , given by (17) has to be considered. $R_k(\beta_0 l)$ and $R_k'(\beta_0 l)$ are approximated as for Case 1; however, since $\beta_1 l \gg 1$, the quantity $R_k(\beta_1 l)$ is given by (13) and S_k is approximated [24] by the first term of the asymptotic expansion for $J_\nu(y)$.

$$S_k(y) = \sqrt{\frac{\pi y}{2}} J_{k+1/2}(y) = \cos \left[y - \frac{(k+1)}{2} \pi \right]. \quad (21)$$

$$|f_1| = \frac{3}{2} \left[-\frac{\eta}{\eta_1} \frac{1}{x_0} - \frac{y^2 \sin y + y \cos y - \sin y}{y(\sin y - y \cos y)} \right]^{-1} \quad (26)$$

$$|f_3| = \frac{7}{(3)4} \left[-\frac{\eta}{\eta_1} \frac{3}{x_0} - \frac{-y^4 \sin y - 6y^3 \cos y + 21y^2 \sin y + 45y \cos y - 45 \sin y}{y(y^3 \cos y - 6y^2 \sin y - 15y \cos y + 15 \sin y)} \right]^{-1}. \quad (27)$$

As a consequence, for $x_0 = \beta_0 l \ll 1$ and $y = \beta_1 l \gg 1$,

$$|f_k| = \frac{2k+1}{k(k+1)} \frac{\eta_1 x_0}{\eta k - \eta_1 x_0 \tan y} \quad (22)$$

$$\left| \frac{f_{k+2}}{f_k} \right| = \frac{(2k+5)k(k+1)}{(k+2)(k+3)(2k+1)} \cdot \frac{\eta(k) - \eta_1 x_0 \tan y}{\eta(k+2) - \eta_1 x_0 \tan y}. \quad (23)$$

It is obvious that $|f_{k+2}/f_k| < 1$, and particularly $|f_3/f_1| < 1$, unless the denominator of (23) approaches zero. In that case $|f_{k+2}/f_k|$ would seem to approach infinity indicating that the radiation field given by (16) does not converge. However, small real parts of R_k and R_k' were neglected in obtaining (22) and (23). If we let

$$\frac{R_k'}{R_k} = -\frac{k}{x_0} + im \quad m \ll \frac{k}{x_0} \quad (24)$$

(22) would have to be modified.

$$|f_k| = \frac{2k+1}{k(k+1)} \frac{\eta_1 x_0}{|\eta k - \eta_1 x_0 \tan y - im\eta_1 x_0|} = \frac{2k+1}{k(k+1)} \frac{x_0}{\left[\left(\frac{\eta}{\eta_1} k - x_0 \tan y \right)^2 + m^2 x_0^2 \right]^{1/2}}. \quad (25)$$

The denominator of this expression can obviously not be zero and consequently $|f_k|$ or $|f_{k+2}/f_k|$ never become infinite. However, the values of y which would make the denominator of (22) zero, still seem to indicate critical points where $|f_{k+2}/f_k|$ becomes larger than one, although not infinite. Critical points of this sort actually do exist, but it turns out, as shown below, that (22) and (23) are very poor approximations for $|f_k|$ and $|f_{k+2}/f_k|$ in the neighborhood of these points. Also, (22) and (23) locate only some of these critical points and these not very accurately.

If we want to show that the antenna pattern (16) significantly deviates from the radiation pattern of a short dipole it is sufficient to show that $|f_3/f_1| > 1$. The functions $P_k'(\cos \theta)$ are plotted for $k=1, 3, 5, 7$ in Figs. 8 to 10. $P_1'(\cos \theta) = -\sin \theta$ is the pattern of the short dipole; if any of the higher coefficients f_k are comparable to f_1 in magnitude, or larger, it is clear that the pattern will not be that of the short dipole.

The exact expressions [26] for $S_k(y)$ and $S_k'(y)$ give

If only the leading terms in the fractions of trigonometric functions in (26) and (27) were considered, these equations would reduce to special cases of (22); however, it is clear that when $y \rightarrow (2n+1)(\pi/2)$ the terms $y^2 \cos y$ and $y^4 \cos y$ in (26) and (27) are zero and any valid approximation must include terms of the next lower order in y if they contain $\sin y$. The following expression is valid for any $y \gg 1$.

$$\left| \frac{f_3}{f_1} \right| = \frac{14}{36} \frac{\left[-\frac{\eta}{\eta_1} \frac{1}{x_0} - \frac{y \sin y + \cos y}{(\sin y - y \cos y)} \right]}{\left[-\frac{\eta}{\eta_1} \frac{3}{x_0} - \frac{-y \sin y - 6 \cos y}{y \cos y - 6 \sin y} \right]}. \quad (28)$$

The critical points where $|f_3/f_1| \gg 1$ are obviously located, where either the numerator approaches infinity or the denominator becomes zero, the conditions are

$$\tan y = y \quad (29)$$

$$\frac{\eta}{\eta_1} \frac{3}{x_0} = \frac{y \sin y + 6 \cos y}{y \cos y - 6 \sin y}. \quad (30)$$

These two equations can be solved graphically as indicated on Figs. 11 and 12.

Some critical values obtained by (29) and (30) were compared with the corresponding values obtained by a digital computer evaluation of (16). The results are as shown in Table I.

TABLE I

| $\frac{y}{x_0} = \frac{\beta_1 l}{\beta_0 l}$ | $\frac{\eta}{\eta_1}$ | By (29) or (30) | By (16) (IBM-650) |
|---|-----------------------|-----------------|-------------------|
| 10 | 1 | 4.49 | 4.49 |
| 10 | 10 | 4.49 | 4.49 |
| 10 | 1 | 7.71 | 7.73 |
| 10 | 10 | 7.71 | 7.73 |
| 30 | 1 | 7.71 | 7.73 |
| 30 | 1 | 7.07 | 6.91 |
| 30 | 1 | 6.91 | 6.76 |

On Figs. 13 and 14 the ratio $|f_3/f_1|$ as obtained by (30) in the neighborhood of a critical point is compared with the results of a digital computer evaluation of the exact coefficients.

As pointed out earlier, whenever $|f_3/f_1| > 1$, the radiation pattern of the antenna will deviate from the dipole pattern (see Fig. 15). In the neighborhood of the critical points some of the higher order coefficients ($k > 3$) are also not negligible. The addition of several of the corresponding associated Legendre function $P_k'(\cos \theta)$ (Figs. 8 to 10) will give patterns which depend upon the relative magnitude and phase of the coefficients. In general these patterns will be multilobed; however, a fortuitous combination of amplitudes and phases leading to a pair of very narrow lobes is not excluded. A small radiator producing such a pattern would be a true "super-gain" antenna exhibiting high directive gain; however, even a small antenna (much smaller than $\lambda/2$) which exhibits a multilobe pattern would be a super-gain antenna in the sense that ordinarily such patterns can only be obtained from radiators which are much longer than half a wavelength.

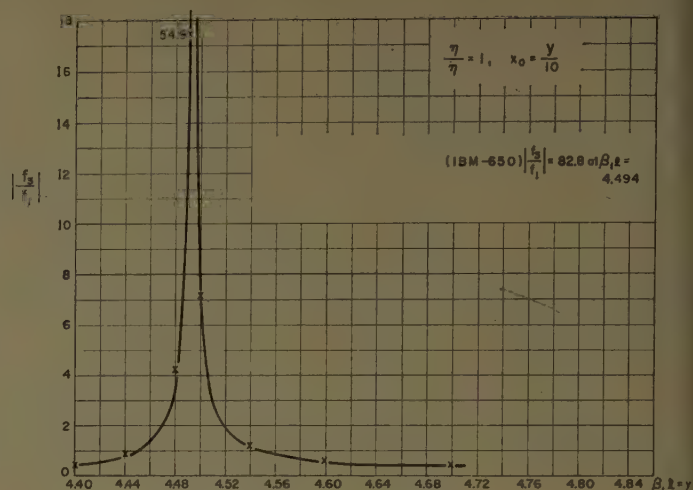


Fig. 13—Comparison of (28) with digital computer evaluation of (17). Solid line represents (17); x represents (28).

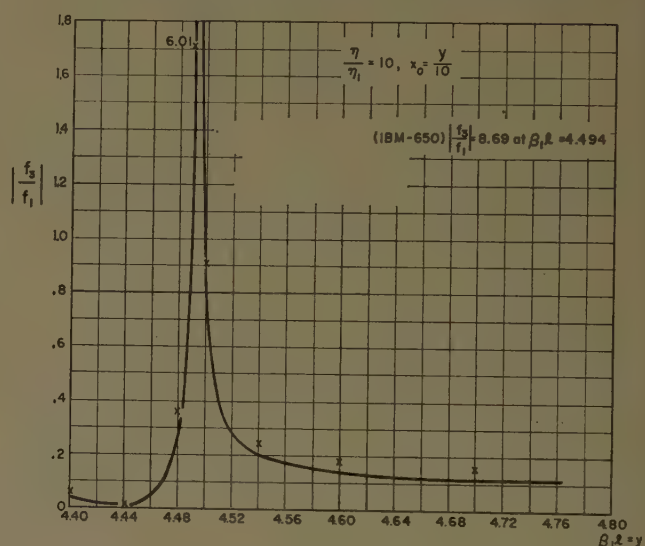


Fig. 14—Comparison of (28) with digital computer evaluation of (17). Solid line represents (17); x represents (28).

$$y = \beta_1 l = 6.965$$

$$\frac{\eta}{\eta_1} = \sqrt{\frac{\epsilon}{\mu}} = 10$$

$$\mu = 1, \epsilon = 100, l = 0.11085 \lambda_0$$

$$\psi = 2^\circ$$

--- SHORT DIPOLE IN AIR

— $y = 6.965$ ANTENNA

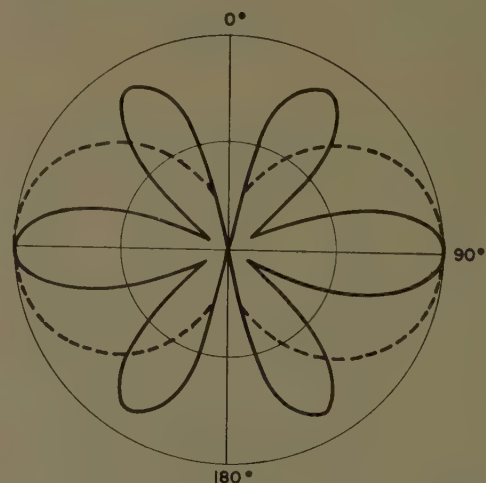


Fig. 15—Normalized radiation pattern of short biconical antenna near a critical frequency.

REFERENCES

- [1] R. E. Burgess, "Iron-cored loop receiving aerial," *Wireless Engineer*, pp. 172-178; June, 1946.
- [2] L. Page, "The magnetic antenna," *Phys. Rev.*, pp. 645-648; June, 1946.
- [3] R. H. Barfield and R. E. Burgess, "Small aerials in dielectric media," *Wireless Engineer*, pp. 246-253; August, 1948.
- [4] W. L. Weeks, "Input Impedance of a Spherical Ferrite Antenna with a Latitudinal Current," Elec. Eng. Res. Lab., University of Illinois, Urbana, Tech. Rept. No. 6; August 20, 1955.
- [5] V. H. Rumsey and W. L. Weeks, "Impedance of Small Ferrite Loop Antennas," Elec. Engrg. Res. Lab., University of Illinois, Urbana, Tech. Rept. No. 13; October 15, 1956.
- [6] W. L. Weeks, "On the Estimation of Ferrite Loop Antenna Impedance," Elec. Engrg. Res. Lab., University of Illinois, Urbana, Tech. Rept. No. 17; April 10, 1957.
- [7] W. Mennerick, "Die Ferrileantenne im Rundfunkgeraet," *Elektrotech. Z.*, vol. 75, pp. 466-468; July 11, 1954.
- [8] J. L. Stewart, "On ferrite loop antenna measurements," 1957. IRE CONVENTION RECORD, p. 46.
- [9] J. L. Stewart, "Research in magnetic antennas," Calif. Inst. of Tech., Pasadena, Contract DA-36-039, SC-73189, ASTIA AD140500; July, 1957.
- [10] O. R. Cruzan, "Radiation Properties of a Spherical Ferrite Antenna," Diamond Ordnance Fuze Lab., Washington, D. C., Tech. Rep. 387; October 15, 1956.
- [11] J. Herman, "Thin Wire Loop and Thin Biconical Antennas in Finite Spherical Media," Diamond Ordnance Fuze Lab., Washington, D. C., Tech. Rept. 462; May 1, 1957.
- [12] O. R. Cruzan, "VHF Ferrite Antenna Radiation Properties," Diamond Ordnance Fuze Lab., Washington, D. C., Tech. Rept. 516; August 15, 1957.
- [13] H. A. Dropkin, E. Metzger and J. C. Cacheris, "VHF Ferrite Antenna Radiation Measurements," Diamond Ordnance Fuze Lab., Washington, D. C., Tech. Rept. 484; July 10, 1957.
- [14] D. M. Grimes, "A Resonant Miniaturized Antenna," Engrg. Res. Inst., The University of Michigan, Ann Arbor, Tech. Memo. No. 45; June, 1957.
- [15] J. D. Kraus, "Antennas," McGraw-Hill Book Co., Inc., New York, N. Y.; 1950.
- [16] S. A. Schelkunoff, "Electromagnetic Waves," D. Van Nostrand Co., Inc., Princeton, N. J., p. 459; 1943.
- [17] H. P. Westman *et al.*, "Reference Data for Radio Engineers," International Telephone and Telegraph Corp., p. 674; 1956.
- [18] W. K. H. Panofsky and M. Phillips, "Classical Electricity and Magnetism," Addison-Wesley Publishing Co., Inc., Cambridge, Mass., pp. 220-227; 1955.
- [19] S. A. Schelkunoff, "Kirchhoff's formula, its vector analogue, and other field equivalence theorems," in "The Theory of Electromagnetic Waves, a Symposium," Interscience Publishers, Inc., New York, N. Y., pp. 107-124; 1951.
- [20] S. A. Schelkunoff, "Theory of antennas of arbitrary size and shape," *Proc. IRE*, vol. 29, pp. 493-521; September, 1941.
- [21] C. T. Tai, "On the theory of biconical antennas," *J. of Appl. Phys.*, vol. 19, pp. 1155-1160; December, 1958.
- [22] S. A. Schelkunoff, "Advanced Antenna Theory," John Wiley and Sons, Inc., New York, N. Y., pp. 43, 45, 54, 68; 1952.
- [23] C. T. Tai, "On Radiation and Radiating Systems in the Presence of a Dissipative Medium," Cruft Lab., Harvard University, Cambridge, Mass., Tech. Rept. No. 77; May 10, 1949.
- [24] S. A. Schelkunoff, "Applied Mathematics for Engineers and Scientists," D. Van Nostrand Co., Inc., Princeton, N. J.; 1948.
- [25] P. M. Morse and H. Feshbach, "Methods of Theoretical Physics," McGraw-Hill Book Co., Inc., New York, N. Y.; 1953.
- [26] I. N. Sneddon, "Special Functions of Mathematical Physics and Chemistry," Oliver and Boyd, London, England, and Interscience Publishers, New York, N. Y.; 1956.
- [27] "Immittance of Thin Biconical Antennas Containing Material of Arbitrary Permeability and Permittivity," RCA Laboratories, Digital Computer Problem No. 619; January, 1958.
- [28] H. Kulsrud, "A Set of Interpretive Subroutines for Cylindrical and Spherical Bessel Functions of the First and Second Kind and Their Derivatives," RCA Labs., internal report; February 25, 1958.
- [29] F. E. Terman, "Radio Engineering," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 54-56; 1937.
- [30] E. C. Jordan, "Electromagnetic Waves and Radiating Systems," Prentice-Hall, Inc., New York, N. Y., p. 460; 1950.
- [31] E. Jahnke and F. E. Ende, "Tables of Functions with Formulae and Curves," Dover Publications, New York, N. Y.; 1945.
- [32] L. J. Chu, "Physical limitations of omni-directional antennas," *J. Appl. Phys.*, vol. 19, pp. 1163-1175; December, 1946.
- [33] H. L. Knudsen, "The field radiated by a ring quasi-array of an infinite number of tangential or radial dipoles," *Proc. IRE*, vol. 41, pp. 781-789; June, 1953.

The Calculated Phase Velocity of Long End-Fire Uniform Dipole Arrays*

F. SERRACCHIOLI† AND C. A. LEVIS‡

Summary—Surface waves have been applied recently to the study of end-fire arrays.¹⁻⁴ An important step in the theory of surface-wave antennas is the calculation of the local phase velocity of the surface wave from the antenna geometry. While experimental work has been done along this line,⁴⁻⁶ no adequate theoretical solution seems to be available.

In this report, a mathematical procedure is presented for obtaining an approximate value of the phase velocity for long end-fire uniform dipole arrays, as a function of the antenna geometry. Results have been reported for a number of geometrical parameters and curves plotted to facilitate the analysis and design of arrays. The results have been compared with experimental measurements and an explanation has been given for the discrepancies found; finally, a criterion for correcting the results has been outlined.

I. INTRODUCTION

CONSIDER the uniform array of the perfectly conducting dipoles shown in Fig. 1. We may think of these as center-fed dipoles, short-circuited at the terminals. The terminal voltage of each element, which is zero, can be obtained by superposition from the voltage produced by the primary source plus the voltages induced by the currents in all the elements.

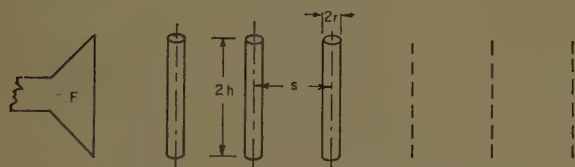


Fig. 1—Uniform array of parasitic dipoles.

* The research reported herein was supported in part by the U. S. Army Signal Research and Development Laboratory, Fort Monmouth, N. J.

† Antenna Lab., Dept. of Elec. Engrg., The Ohio State University, Columbus, Ohio. Visiting research scientist by appointment, supported by the International Cooperation Administration under a program administered by the National Academy of Sciences.

‡ Antenna Lab., Dept. of Elec. Engrg., The Ohio State University, Columbus, Ohio.

¹ J. C. Simon and G. Weill, "Sur le rayonnement longitudinal d'antennes diélectriques," *Compt. rend.*, vol. 235, pp. 1379-1381; December, 1952.

² J. C. Simon and G. Weill, "Un nouveau type d'aérien à rayonnement longitudinal," *Ann. Radioélectricité*, vol. 8, pp. 183-193; July, 1953.

³ J. C. Simon and V. Biggi, "Un nouveau type d'aérien et son application à la transmission de télévision à grande distance," *L'Onde Electrique*, vol. 34, pp. 883-896; November, 1954.

⁴ H. W. Ehrenspeck and H. Poehler, "A New Method for Obtaining Maximum Gain from Yagi Antennas," Electronics Res. Directorate, AFRC, Bedford, Mass., Rept. ERD CRRDA-TM-56-123; August, 1956.

⁵ R. G. Malech, "Stripline radiators," 1955 IRE CONVENTION RECORD, pt. 1, p. 51.

⁶ D. K. Reynolds, "Broadband traveling wave antennas," 1957 IRE NATIONAL CONVENTION RECORD pt. 1, pp. 99-107.

Thus,

$$V_0 = V_0(F) + \sum_n I_n Z_{0n} = 0, \quad (1)$$

where V_0 is the terminal voltage at the general zeroth element, $V_0(F)$ is the direct contribution of the source, I_n is the terminal current of the element n , and Z_{0n} is the mutual impedance between elements 0 and n . A similar equation can be written for each element, and, in principle, the resulting set can be solved for the currents I_n , but no adequate method of handling the numerical work exists when many elements are used. We will assume here that the array is very long, that the zeroth element is somewhere in its middle, and that the geometry is such that a "trapped" wave can be set up and supported by the structure. Because at a sufficiently large distance from the source the guided-wave field is presumably much larger than the source field, the latter can be neglected.

If the array is infinitely long, (1) becomes

$$V_0 = \sum_{-\infty}^{+\infty} Z_{0n} I_n = 0, \quad (2)$$

and the guided wave concept leads to the assumption that the terminal currents I_n must be of the form

$$I_n = I_0 e^{-n\gamma s}, \quad (3)$$

where $\gamma = \alpha + j\beta$ is the unknown propagation constant and s is the distance between adjacent elements in the array, as illustrated in Fig. 1. If a trapped wave exists, then $\alpha = 0$. (For a detailed discussion, see Appendix III.)

Upon substituting (3) into (2) and dividing through by I_0 , the following equation results:

$$\sum_{-\infty}^{+\infty} Z_{0n} e^{-in\beta s} = 0. \quad (4)$$

The following section is devoted to finding a suitable method for solving this fundamental equation.

II. TRANSFORMATION OF THE FUNDAMENTAL EQUATION

The bilateral series on the left side of (4) can be written in the form of a unilateral trigonometric series, simply by combining the terms symmetrically with respect to $n = 0$. Thus,

$$Z_{00} + 2 \sum_{n=1}^{\infty} Z_{0n} \cos(n\beta s) = 0. \quad (5)$$

The same series may also be expressed as the sum of two unilateral power series by letting

$$x = e^{-i\beta s}. \quad (6)$$

In this way, (4) becomes

$$Z_{00} + \sum_{n=1}^{\infty} Z_{0n} x^n + \sum_{n=1}^{\infty} Z_{0n} \frac{1}{x^n} = 0. \quad (7)$$

The mutual impedances Z_{0n} , even after some simplifying approximations discussed in Section V, are rather complicated functions of the array geometry (see Appendix I); however, as n grows very large their asymptotic expression is particularly simple. Physical considerations lead immediately to the statement that, when $n\beta_0 s \gg 1$, β_0 being the propagation constant of free space, Z_{0n} must be of the form

$$Z_{0n} \simeq \frac{jc}{\beta_0 s} \frac{e^{-in\beta_0 s}}{n}, \quad (8)$$

where c is a real constant depending on the dipole geometry; its expression can be found in Appendix II. Let

$$k = \frac{c}{\beta_0 s}, \quad x_0 = e^{-i\beta_0 s}, \quad (9)$$

and consider the two series:

$$S(x, x_0) \equiv jk \sum_{n=1}^{\infty} \frac{(x_0 x)^n}{n} \quad (10)$$

$$S'(x, x_0) \equiv jk \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{x_0}{x} \right)^n. \quad (11)$$

Since x and x_0 both have modulus 1, S and S' are convergent, provided that $x \neq x_0^{-1}$ and $x \neq x_0$, respectively (see Appendix II), and have the sums:

$$S(x, x_0) \equiv -jk \log(1 - x_0 x), \quad [x \neq x_0^{-1}] \quad (12) \quad \text{or}$$

$$S'(x, x_0) \equiv -jk \log \left(1 - \frac{x_0}{x} \right), \quad [x \neq x_0] \quad (13)$$

$$S + S' \equiv -jk \log \left[\left(1 - x_0 x \right) \left(1 - \frac{x_0}{x} \right) \right], \quad [x_0 \neq x \neq x_0^{-1}]. \quad (14)$$

If $(S + S')$ is subtracted from both sides of (7), with the notation

$$Z_{0n} - jk \frac{x_0^n}{n} = A_n, \quad (15)$$

(7) assumes the form

$$Z_{00} + \sum_{n=1}^{\infty} A_n x^n + \sum_{n=1}^{\infty} A_n \frac{1}{x^n} = jk \log \left[(1 - x_0 x) \left(1 - \frac{x_0}{x} \right) \right], \quad (16)$$

or else, by (6) and (9),

$$Z_{00} + 2 \sum_{n=1}^{\infty} A_n \cos(n\beta s) - jk \log(1 + e^{2i\beta_0 s} - 2e^{-i\beta_0 s} \cos \beta s) = 0. \quad (17)$$

The fundamental advantage of using (17) instead of (5) is that the series

$$\sum_{n=1}^{\infty} A_n \cos(n\beta s) \quad (18)$$

is more rapidly convergent than the series on the left side of (5). In Appendix III it is shown that the coefficients A_n decrease at least as $1/n^2$, while the Z_{0n} decrease as $1/n$. Furthermore, even if the series of (18) is neglected, (17) approximates an infinite array better than the series of (5) when the summation is (necessarily) truncated at a finite n . This was observed from computational checks.

We can note immediately, from (2) and all the following, that for every solution β , another solution $-\beta$ exists; the former represents a wave traveling from left to right, the latter a wave from right to left, the two waves being perfectly symmetrical.

From physical evidence it is known that a trapped wave is a slow wave, and consequently we can put the following limitations on β for the surface-wave mode under consideration:

$$\beta_0 < \beta \leq \frac{\pi}{s} \quad (\beta > 0) \quad (19)$$

$$-\frac{\pi}{s} \leq \beta < -\beta_0 \quad (\beta < 0). \quad (20)$$

III. THE SOLUTION OF THE TRANSFORMED EQUATION

Eq. (17) is equivalent to two real equations, one for the real and one for the imaginary parts. Noting that

$$\log(1 + e^{-2i\beta_0 s} - 2e^{-i\beta_0 s} \cos \beta s) = \log[2(\cos \beta_0 s - \cos \beta s)] - j\beta_0 s, \quad (21)$$

from (17) we get:

$$R_{00} + 2 \sum_1^{\infty} R[A_n] \cos(n\beta s) - k\beta_0 s = 0 \quad (22)$$

$$X_{00} + 2 \sum_1^{\infty} I[A_n] \cos(n\beta s) - k \log [2(\cos \beta_0 s - \cos \beta s)] = 0. \quad (23)$$

The propagation constant for an actual array, described by a linear system of equations of the type in (1), one for each element of the array, exists and is unique. Since the attenuation constant α has been neglected, β must consequently satisfy both (22) and (23) simultaneously over a finite interval. In fact, with the approximate formulas used for the impedances, only the self-reactance X_{00} depends on the dipole diameter; that is, when $h < \lambda/4$, the self-reactance is negative with a logarithmic singularity at $r=0$. Since the series term in (23) is everywhere continuous and bounded, we have from (23)

$$\lim_{r \rightarrow 0} \beta = \pm \beta_0. \quad (24)$$

In order for the preceding to be true, (22) must not be independent of (23) over the interval (19) or (20). Practical considerations connected with the numerical solution of (23) further restrict these limits in some cases.

The computations carried out have shown that the above condition is verified quite satisfactorily in all cases; the variation of β , introduced by allowing an $\alpha \neq 0$ in (3), is always less than 0.01 per cent.

The problem of finding the relative phase velocity $v/c = \beta_0/\beta$ has then been reduced to solving (23) for β . The method used in solving this equation is the Newton-Raphson method.⁷ The function $X(\beta s)$ representing the left side of (23) is regular for all values of β except $\beta = \pm \beta_0$; hence an approach to the null value of $X(\beta s)$ may be attempted by using the first two terms in a Taylor's series expansion about a selected value of β , $\beta^{(0)}$. In this way an increment $\delta\beta$ is found and the procedure is repeated, carrying out the expansion about the new value $\beta^{(1)} = (\beta^{(0)} + \delta\beta)$, and so on. Thus, the formula for the successive increments is found to be

$$\delta(\beta s) = - \frac{X(\beta^{(i)} s)}{X'(\beta s) |_{\beta = \beta^{(i)}}}, \quad (25)$$

where $X'(\beta s)$ denotes the derivative of $X(\beta s)$ with respect to its argument, and β^i are successive values of β obtained by carrying out i iterations.

To the extent that the sign of $X'(\beta s)$ does not change (this is examined in more detail in the next section), the root of (23) is ultimately approached, no matter what $\beta^{(0)}$ we start from.

⁷ K. L. Nielsen, "Methods in Numerical Analysis", The Macmillan Co., New York, N. Y.; 1956.

IV. DISCUSSION OF THE SOLUTION

Consider (23) written in the following way:

$$k \log [2(\cos \beta_0 s - \cos \beta s)] - 2 \sum_1^{\infty} I[A_n] \cos(n\beta s) = X_{00}, \quad (26)$$

and the derivative of the left side with respect to (βs) :

$$k \frac{\sin \beta s}{\cos \beta_0 s - \cos \beta s} + 2 \sum_1^{\infty} n I[A_n] \sin(n\beta s). \quad (27)$$

As was pointed out before, the coefficients A_n decrease in magnitude at least as $1/n^2$; practical cases show that their phase varies ultimately as $e^{-i\beta_0 n}$, so that at most a logarithmic singularity may be found in the series term of (27) at $\beta = \beta_0$. However, the first term of (27) has a pole of the first order at $\beta = \beta_0$; hence, in any case the derivative (27) approaches $+\infty$ as $\beta \rightarrow \beta_0 +$. The same derivative is zero when $\beta = \pi/s$.

Fig. 2 shows a possible behavior of the left member of (26). Curve 1 is the logarithmic term and curve 2 is the result of its sum with the series.

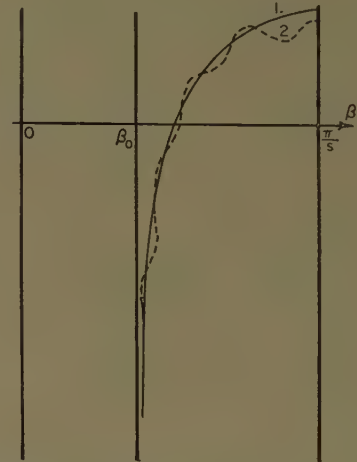


Fig. 2—Qualitative behavior of the functions $f_1 = K \log 2 (\cos \beta_0 s - \cos \beta s)$ (curve 1); $f_2 = f_1 - 2 \sum_1^{\infty} I[A_n] \cos(n\beta s)$ (curve 2).

It is apparent that a negative derivative may be found only in the region of very low phase velocity; that is, not in the region of interest. Numerical computations have shown that the derivative (27) is actually positive over the range of interest; furthermore, with sufficiently small diameters (the upper limit is set by the assumption of sinusoidal current distribution), even with a curve like curve 2 of Fig. 2, the solution is unique.

We conclude that, within practical limits, the solution of (23) exists and is unique. Eq. (23), in connection with (24) and Fig. 2, shows also a qualitative behavior of v/c . For $h < 0.25 \lambda$, the self-reactance X_{00} is negative and decreases in magnitude with increasing h and/or r . Therefore:

- 1) For a given s , increasing h lowers v/c ;
- 2) Increasing s narrows the interval $(\beta_0, \pi/s)$ and the curve of Fig. 2 remains closer to $\beta=\beta_0$ up to a certain point. Then it sharply departs from $\beta=\beta_0$ to end with zero derivative at $\beta=\pi/s$. This same behavior will be shown qualitatively by a curve of v/c vs h/λ or r/λ : the larger s/λ , the closer to 1 will v/c be at first; then, at a certain value of h/λ or r/λ the curve drops sharply and may cross another curve with smaller s/λ .

V. NUMERICAL COMPUTATIONS

In order to construct tables of the relative phase velocity v/c as a function of array parameters, (23) has been solved for a number of different array configurations in a range of practical interest.

Extensive use was made of an IBM 650-653 B4 digital electronic computer, and for this purpose it was necessary to convert (23) into a form that could be conveniently programmed for computer use.

The values of the mutual and self-impedances were calculated from the formulas of Brown and King,⁸ which lend themselves to automatic computation. The fundamental assumptions upon which they are based are: 1) the current distribution is sinusoidal over the dipole length, with zero magnitude at the ends; 2) the thickness of the dipoles is neglected in the mutual impedance calculations; 3) in the computation of the self-impedance of a dipole and of the mutual impedance between any two dipoles, the effect of the remaining array is ignored. These assumptions should be reasonably well approximated for the dipole dimensions and spacings used in common end-fire arrays.

More accurate formulas could be used,⁹ but the labor involved in programing these for automatic computation is prohibitive. Even with the simplified formulas of Brown and King,⁸ the expressions for the $Z_{0n} = R_{0n} + jX_{0n}$ are rather complicated, as may be seen in Appendix I.

Another approximation comes from the necessity of truncating the series in (23) at a finite N . The empirical formula used for N was

$$N = 17 \cdot \frac{h}{s}, \quad (28)$$

rounded to the first higher integer.

VI. RESULTS OF NUMERICAL COMPUTATIONS

The parameters used are reported in Table I; their meaning is illustrated in Fig. 1.

⁸ G. H. Brown and R. King, "High-frequency models in antenna investigations," *PROC. IRE*, vol. 22, pp. 957-980; April, 1934. See also, S. A. Schellkunoff, "Electromagnetic Waves," D. Van Nostrand Co., Inc., New York, N. Y., pp. 369-374; 1943.

⁹ S. Uda and Y. Mushiaki, "Yagi-Uda Antenna," *Res. Inst. of Elec. Commun., Tohoku University, Sendai, Japan*; 1954.

TABLE I
CHARACTERISTIC PARAMETERS USED IN THE COMPUTATIONS

| | |
|--------------------------------------|---------------------------|
| $s/\lambda=0.10,$ | $h/s=1.00$ (0.05) 1.70 |
| $s/\lambda=0.125,$ | $h/s=1.00$ (0.05) 1.60 |
| $s/\lambda=0.15,$ | $h/s=0.75$ (0.05) 1.30 |
| $s/\lambda=0.175,$ | $h/s=0.76$ (0.04) 1.12 |
| $s/\lambda=0.20,$ | $h/s=0.50$ (0.05) 1.00 |
| $s/\lambda=0.225,$ | $h/s=0.64$ (0.04) 0.96 |
| $s/\lambda=0.25,$ | $h/s=0.40$ (0.04) 0.84 |
| $s/\lambda=0.275,$ | $h/s=0.55$ (0.025) 0.775 |
| $s/\lambda=0.30,$ | $h/s=0.350$ (0.025) 0.725 |
| $s/\lambda=0.325,$ | $h/s=0.45$ (0.025) 0.675 |
| $s/\lambda=0.35,$ | $h/s=0.300$ (0.025) 0.625 |
| $s/\lambda=0.375,$ | $h/s=0.40$ (0.02) 0.58 |
| $s/\lambda=0.40,$ | $h/s=0.30$ (0.02) 0.52 |
| $r/h=0.1, 0.08, 0.05, 0.025, 0.0125$ | |

In addition to the five values of r/h included in Table I, the value $r/\lambda=0.024$ was also used for one of the parameters, in order to compare the results with the experimental curves reported by Ehrenspeck and Poehler,⁴ who had used this parameter in their measurements. These curves served as a check throughout the computations, as they are the most complete ones available on this subject to date.

In a few cases, intermediate steps of h/s were added to those indicated in Table I, to obtain a better definition of the curves. Curves of v/c for $r/\lambda=0.024$ are reported in Fig. 3, together with the experimental curves measured by Ehrenspeck and Poehler. The curves for the parameters listed in Table I are reported in Fig. 4.

The dependence of v/c on the frequency for a given array may be considered as a measure of the array bandwidth. An approximate estimate of it can be deduced from the results reported; it is necessary only to hold h/s and r/h constant and pick out values of v/c for different s/λ .

In Fig. 5 the dependence of phase velocity on frequency is shown for the five values of r/h used, in an array with $h/s=0.5$. The points that determine the curve are rather widely spaced; a more complete description would require smaller steps in the values of s/λ .

VII. DISCUSSION OF THE RESULTS

It is apparent from Fig. 3 that the calculated values of v/c are uniformly slightly higher than the measured values. An error in the same direction seems to be present in the curves of Fig. 4. It is believed that the approximate formulas used for the impedances account for most of these discrepancies. In this respect, the most important factor is the error which affects the self-reactance X_{00} , because the approximations mentioned in Section V, particularly assumption 1), primarily affect the self-impedance; furthermore, it is probable that the error affecting the mutual impedances is partly compensated for in the summation.

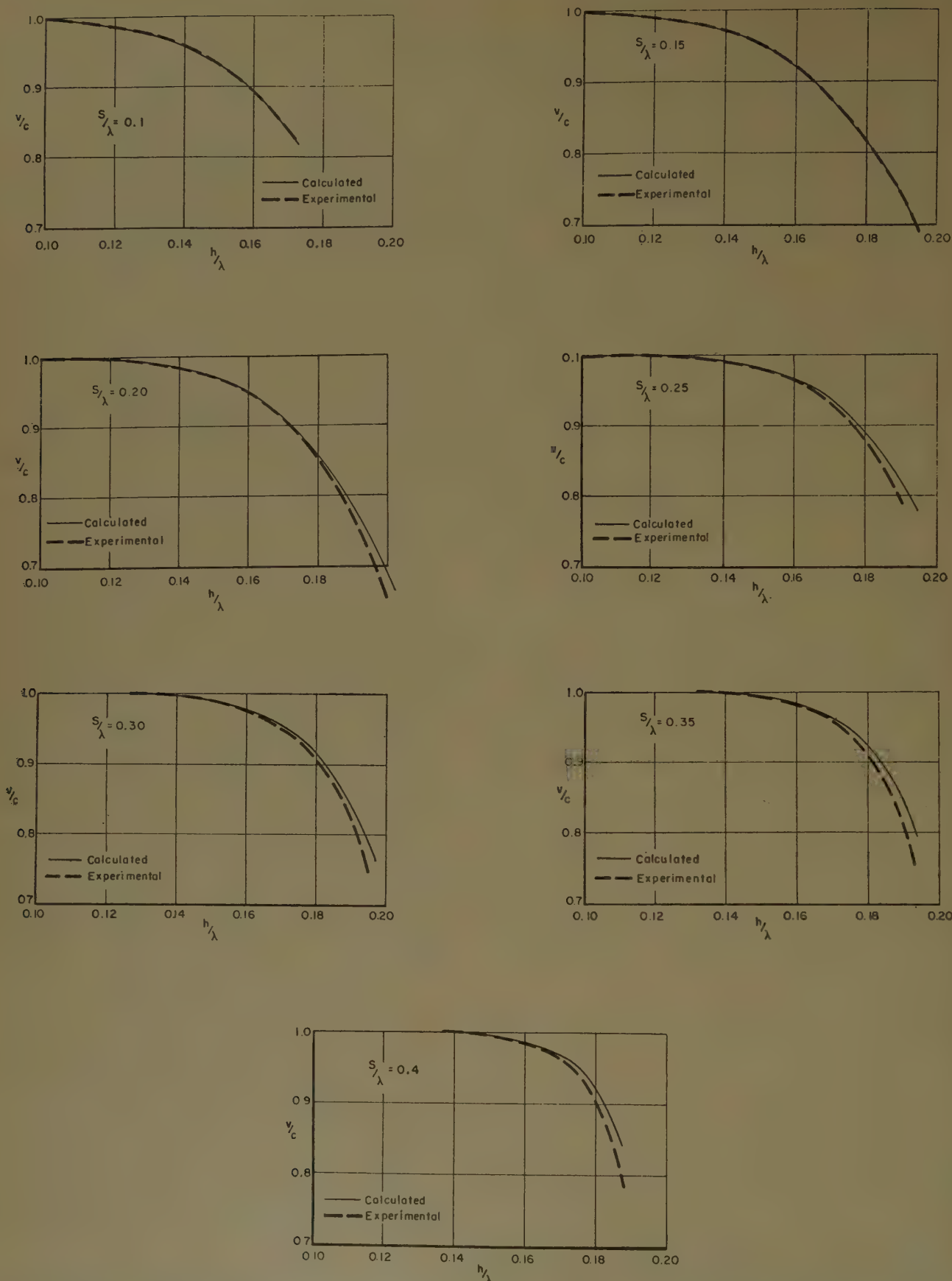


Fig. 3—Phase velocity, relative to free space, on long uniform array of dipoles; solid curves are calculated, the dashed curves were measured by Ehrenspeck and Poehler, h denotes element half-length, s center-to-center spacing, and the radius of all elements is 0.024λ .

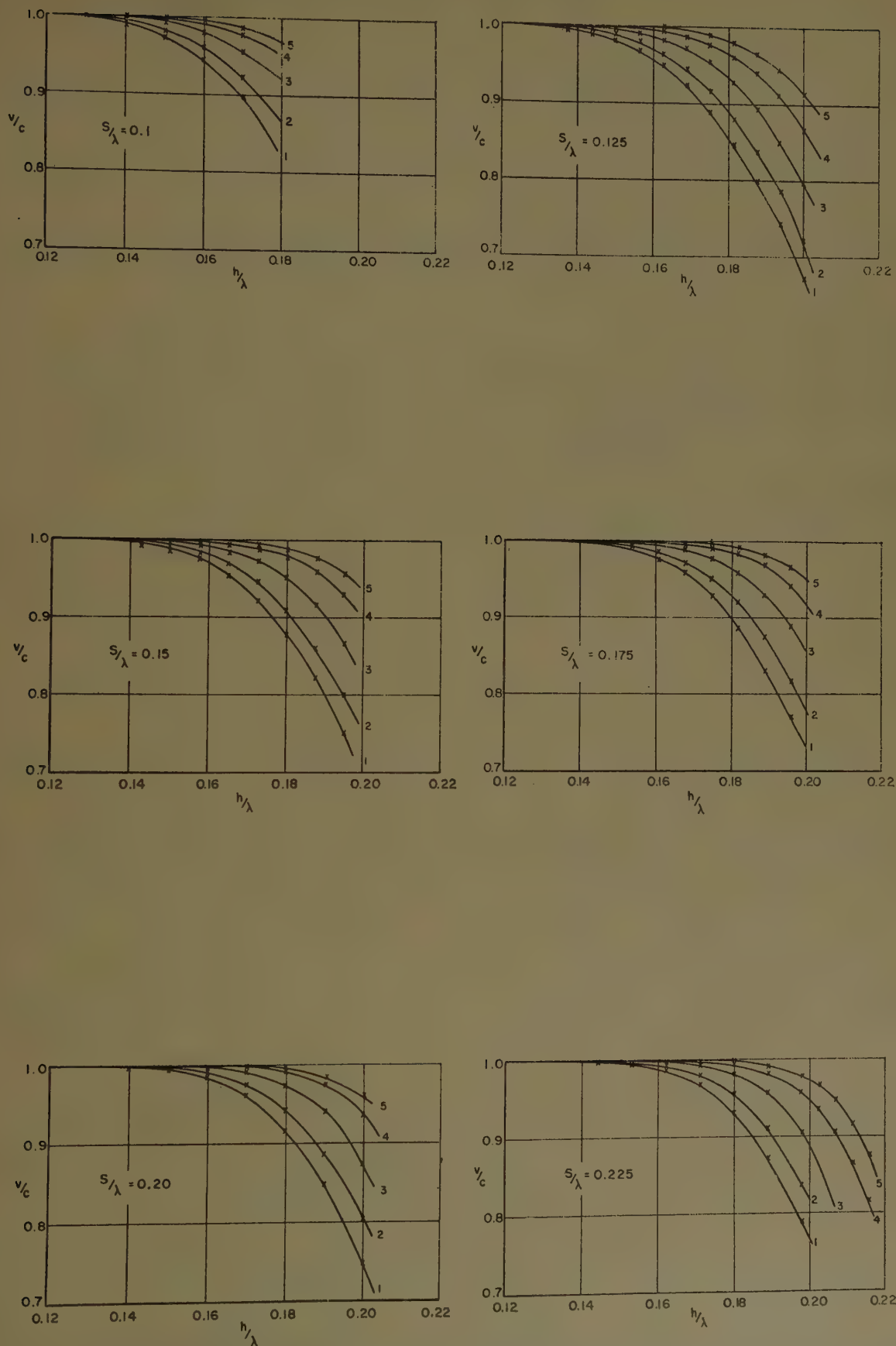


Fig. 4—Calculated phase velocity, relative to free space, on long uniform arrays of dipoles. Curves marked 1 are for $r=0.1h$, 2 for $r=0.08h$, 3 for $r=0.05h$, 4 for $r=0.025h$, 5 for $r=0.0125h$. (Cont'd. on next page.)

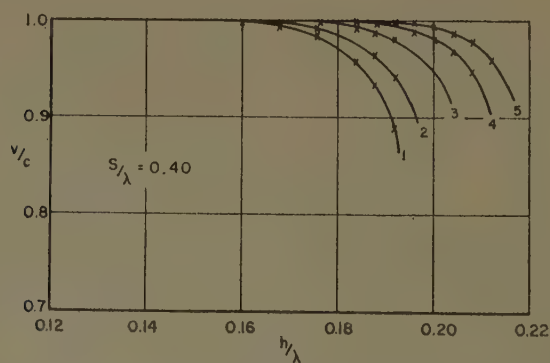
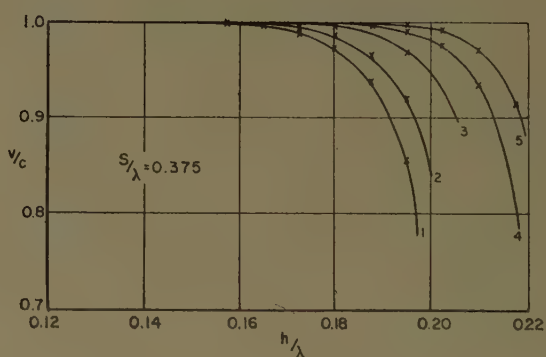
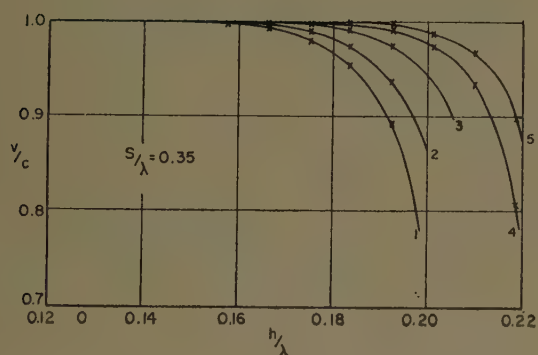
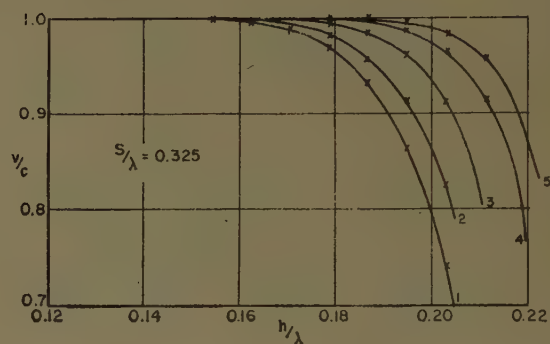
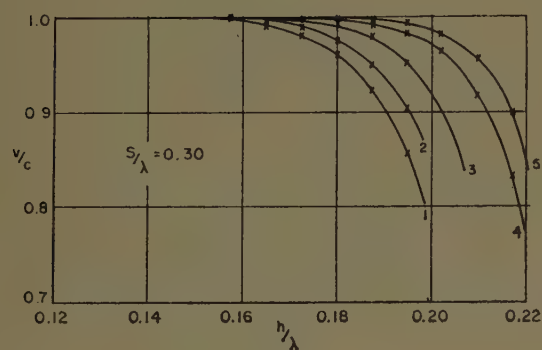
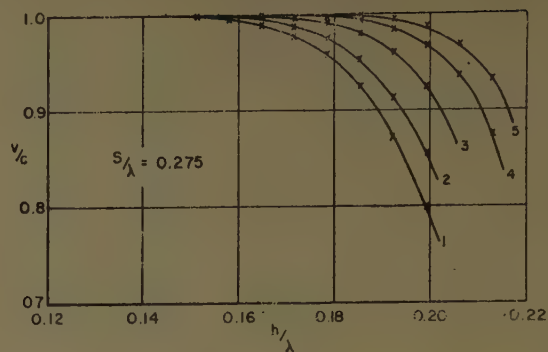
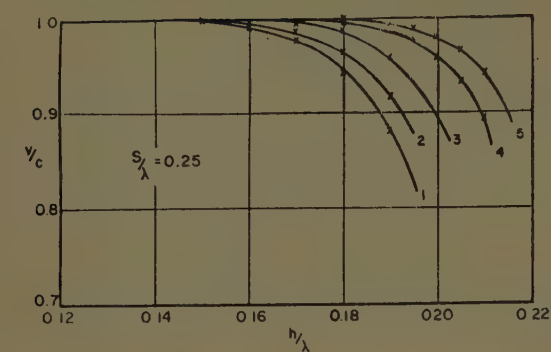


Fig. 4 (Cont'd)—Calculated phase velocity, relative to free space, on long uniform arrays of dipoles. Curves marked 1 are for $r=0.1h$, 2 for $r=0.01h$, 3 for $r=0.05h$, 4 for $r=0.025h$, 5 for $r=0.0125h$.

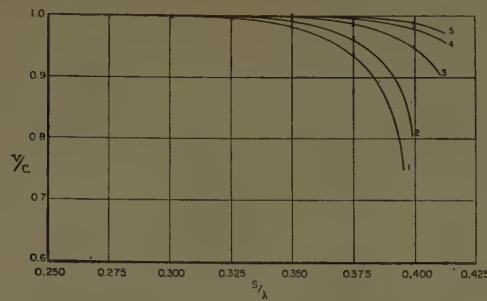


Fig. 5—Calculated phase velocity, relative to free space, as a function of frequency for typical end-fire antennas; $h/s=0.5$ and the radii correspond to those of Fig. 4.

To evaluate the error introduced by these approximations, a comparison is made in Table II between some values of X_{00} calculated with the formula (35) of Appendix I, commonly referred to as the improved zeroth-order approximation, and the corresponding values obtained by R. W. P. King,¹⁰ by using a second-order approximation.

TABLE II

INPUT REACTANCES ACCORDING TO DIFFERENT APPROXIMATIONS.
 $X_{00}^{(0)}$ IS A ZERO-TH-ORDER APPROXIMATION, $X_{00}^{(2)}$ IS A SECOND-ORDER APPROXIMATION

| | $r/h=0.0604$ | | $r/h=0.0133$ | |
|--------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | $X_{00}^{(0)}$ (ohms) | $X_{00}^{(2)}$ (ohms) | $X_{00}^{(0)}$ (ohms) | $X_{00}^{(2)}$ (ohms) |
| $h/\lambda=0.207$ | -32.1 | -30 | -82.4 | -79.7 |
| $h/\lambda=0.19$ | -62.1 | -58 | -134.1 | -127.1 |
| $h/\lambda=0.175$ | -90.0 | -85.6 | -182.5 | -177.4 |
| $h/\lambda=0.143$ | -156 | -151 | -300 | -294 |
| $h/\lambda=0.1114$ | -246 | -237 | -460 | -451 |

Assuming that $X_{00}^{(2)}$ is exact and that the error in the phase velocity depends only on $\Delta X_{00} = (X_{00}^{(0)} - X_{00}^{(2)})$, from (25) we get a first-order correction for β :

$$\Delta(\beta s) = -\frac{\Delta X_{00}}{X'(\beta s)}, \quad (29)$$

where $X'(\beta s)$ is given by (27). Consequently,

$$\frac{v}{c} + \Delta\left(\frac{v}{c}\right) = \frac{\beta_0}{\beta + \Delta\beta} \simeq \frac{\beta_0}{\beta} \left(1 - \frac{\Delta\beta}{\beta_0}\right) \quad (30)$$

$$\Delta\left(\frac{v}{c}\right) \simeq \frac{1}{\beta s} \frac{\Delta X_{00}}{X'(\beta s)}. \quad (31)$$

Since ΔX_{00} is negative and $X'(\beta s)$ is positive, $\Delta(v/c)$ is negative; that is, the value of v/c is lowered if a better approximation for X_{00} is used. The error is more evident when $X'(\beta s)$ is small; that is, for lower phase velocities. Numerical computations show that $\Delta(v/c)$ may range from zero (where $X'(\beta s) \rightarrow \infty$) up to about 4 per cent in the range of interest.

Furthermore, the $X_{00}^{(2)}$ are themselves approxima-

tions, and measurements¹¹ carried out at the Antenna Laboratory of the Ohio State University, Columbus, point to absolute values of X_{00} which are smaller than those given by King. Such a difference is illustrated in Fig. 6, where the echo areas σ for two dipoles are shown. The curves were calculated from a solution obtained by the variational method, which in the range $h/\lambda \leq 0.25$ gives results practically coincident with those obtained by a second-order approximation. The relationship between σ and Z_{00} is

$$\sigma \simeq \frac{W \tan^4\left(\beta_0 \frac{h}{2}\right)}{|Z_{00}|^2}, \quad \frac{h}{\lambda} \leq 0.25, \quad (32)$$

where W is a constant. To the extent that the following relationship is true:

$$R_{00} \frac{\partial(R_{00})}{\partial h} \ll X_{00} \frac{\partial(X_{00})}{\partial h}, \quad (33)$$

the influence of the self resistance may be neglected in the curves of Fig. 6, and the differences between calculated and measured values may be attributed to X_{00} . This is generally the case for moderately thin dipoles.

Furthermore, the difference between theoretical and experimental curves in Fig. 6, in the range $h/\lambda < 0.25$, consists of a rigid translation in the direction of the h/λ axis by a quantity proportional to r/λ ; therefore the same translation should be applied to the curves of v/c given in the text. A translation of r/λ does, in fact, give a better agreement with measurements for moderately thin dipoles. This no longer holds for thick dipoles, such as those referred to in Fig. 3, because in this case the correction of h/λ that brings $X_{00}^{(0)}$ to its correct value is considerably smaller¹² than r/λ .

In any case, the same correction of h/λ that brings $X_{00}^{(0)}$ to its correct measured value may be applied to the results of this report.

¹¹ R. G. Kouyoumjian, "The Calculation of Echo Areas of Perfectly Conducting Objects by the Variational Method," Graduate School of The Ohio State University, Columbus; abstract of Ph.D. thesis; 1953.

¹² F. Serracchioli, "Calculated Self and Mutual Impedances of Short Parallel Dipoles," Antenna Lab., The Ohio State University Res. Foundation, Columbus, Rept. 662-23, March 24, 1959; prepared under Contract DA 36-039 sc-70174, U. S. Army, Engrg. Lab., Fort Monmouth, N. J.

¹⁰ R. King, "The Theory of Linear Antennas," Harvard University Press, Cambridge, Mass., pp. 168-176; 1956.

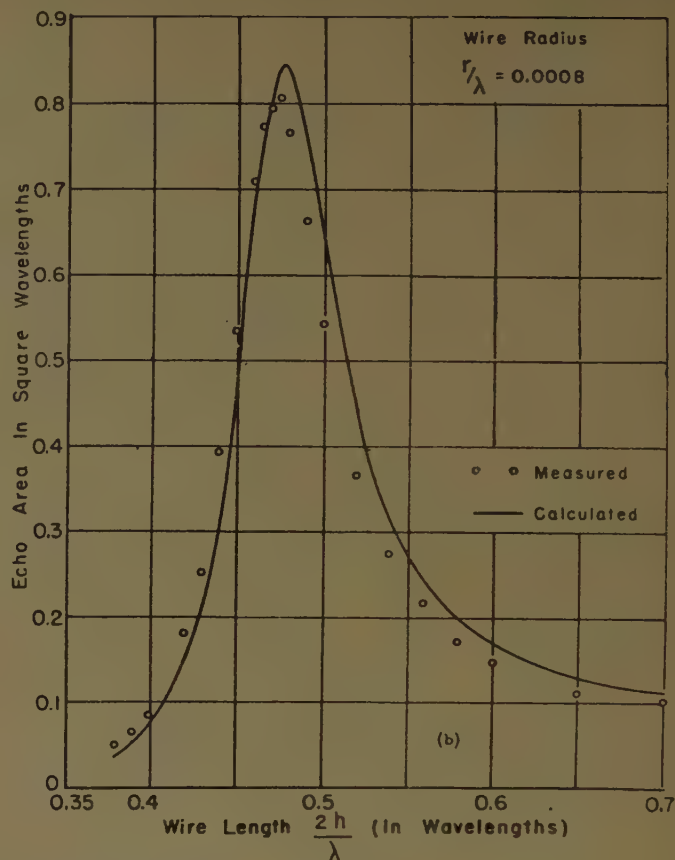
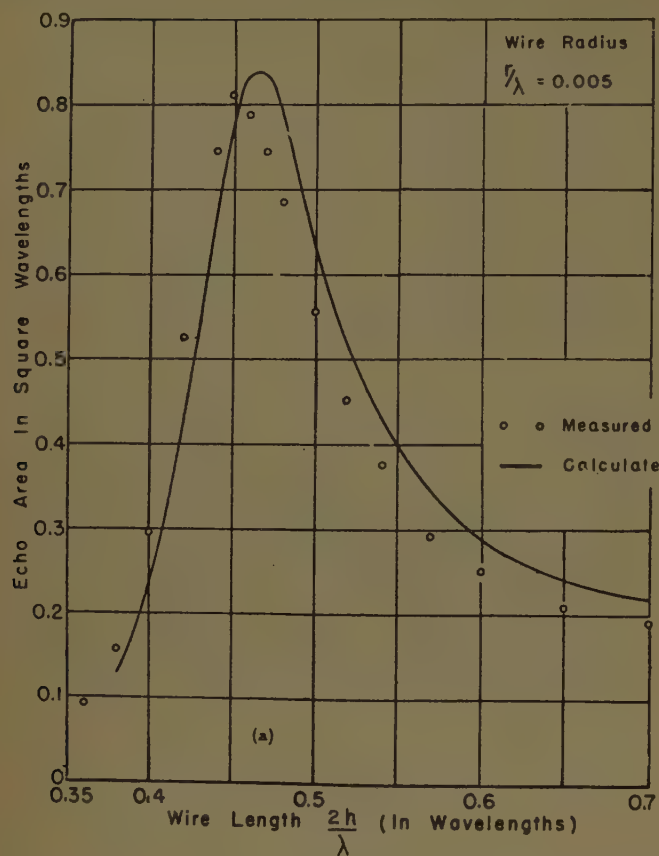


Fig. 6—The echo area of thin straight wires.

VIII. CONCLUSION

The calculation of the phase velocity for long end-fire uniform dipole arrays yields results in good agreement with available experimental data. An empirical correction term for the effective length of the dipoles, similar to that found in previous studies^{11,12} of their self-impedance and back-scattering, will improve this agreement. Such a correction could be expected, since the self-impedances used in the calculations were found by an improved zero-order approximation. If the exact values of the self-impedances are used, the calculated phase velocities are in excellent agreement with experimental data. A correction function for the dipole height may be readily found from a comparison between the exact values of self-reactance and those used in these calculations.¹²

The results of the calculations are given in a form suitable for the design of uniform arrays and the estimation of probable pattern bandwidth directly from the physical dimensions of the arrays. It is also felt that a good approximation of the local phase velocity of modulated arrays may be found from these same results. The results are complete enough to permit a high degree of accuracy in cross-plotting of data for diameters and spacings not calculated.

APPENDIX I

Referring to Fig. 1, and letting

$$G = \beta_0 h,$$

$$D = \beta_0 n s,$$

$$D_+ = \sqrt{D^2 + G^2} + G,$$

$$D_- = \sqrt{D^2 + G^2} - G,$$

$$D_{+2} = \sqrt{D^2 + 4G^2} + 2G, \quad D_{-2} = \sqrt{D^2 + 4G^2} - 2G,$$

and also denoting Euler's constant by

$$C = 0.577215665 \dots,$$

the formulas used for the calculation of the self- and mutual impedances are the following:⁸

$$\begin{aligned} R_{00} = 30 \{ & [1 - \cot^2 G][C + \ln(4G) - Ci(4G)] \\ & + 4 \cot^2 G[C + \ln(2G) - Ci(2G)] \\ & + 2 \cot G[Si(4G) - 2Si(2G)] \}, \end{aligned} \quad (34)$$

$$\begin{aligned} X_{00} = 30 \left\{ 2 \cot G \left[C - \ln \frac{h\lambda}{2\pi r^2} + Ci(4G) - 2Ci(2G) \right] \right. \\ \left. + \cot^2 G[4Si(2G) - Si(4G)] + Si(4G) \right\}, \end{aligned} \quad (35)$$

$$\begin{aligned}
R_{0n} = & 60 \cot^2 G [2 Ci(D) - 2 Ci(D_-) \\
& - 2 Ci(D_+) + Ci(D_{+2}) + Ci(D_{-2})] \\
& + \frac{30}{\sin^2 G} [2 Ci(D) - Ci(D_{-2}) - Ci(D_{+2})] \\
& + 60 \cot G [Si(D_{+2}) - Si(D_{-2}) \\
& - 2 Si(D_+) + 2 Si(D_-)], \quad (36)
\end{aligned}$$

$$\begin{aligned}
X_{0n} = & -60 \cot^2 G [2 Si(D) - 2 Si(D_-) \\
& - 2 Si(D_+) + Si(D_{+2}) + Si(D_{-2})] \\
& - \frac{30}{\sin^2 G} [2 Si(D) - Si(D_{-2}) - Si(D_{+2})] \\
& + 60 \cot G [Ci(D_{+2}) - Ci(D_{-2}) \\
& - 2 Ci(D_+) + 2 Ci(D_-)]. \quad (37)
\end{aligned}$$

The problems involved in the automatic numerical computations of these expressions and the devices adopted to solve them are discussed in detail in a separate technical report.¹²

APPENDIX II

Consider the two following approximate relationships which are valid when $x \gg \delta$ and $x \gg 1$.

$$Si(x + \delta) - Si(x) \simeq \frac{1}{x} [\cos x - \cos(x + \delta)], \quad (38)$$

$$Ci(x + \delta) - Ci(x) \simeq \frac{1}{x} [\sin(x + \delta) - \sin x]. \quad (39)$$

By this approximation, terms of the order of δ/x are neglected with respect to unity. With the same approximation we can write:

$$\sin(\sqrt{x^2 + \delta^2}) \simeq \sin x \quad (40)$$

$$\cos(\sqrt{x^2 + \delta^2}) \simeq \cos x. \quad (41)$$

Introducing the approximations (38), (39), (40) and (41) into formulas (36) and (37), and simplifying we get, when $D \gg G$,

$$R_{0n} \simeq \frac{120}{\sin^2 G} (1 - \cos G)^2 \frac{\sin D}{D}, \quad (42)$$

$$X_{0n} \simeq \frac{120}{\sin^2 G} (1 - \cos G)^2 \frac{\cos D}{D}. \quad (43)$$

Combining the two formulas together:

$$Z_{0n} \simeq j \frac{120}{\beta_0 S} \left(\frac{1 - \cos \beta_0 h}{\sin \beta_0 h} \right)^2 \frac{e^{-jn\beta_0 s}}{n}. \quad (44)$$

Thus, the difference between the formulas (36) and (37) for Z_{0n} and its asymptotic expression (44) is a zero at least of the order of $1/n^2$ as $n \rightarrow \infty$. (It may be of

higher order if the second-order terms cancel out in the difference.) From (44) we also get the value of c in (8):

$$c = 120 \left(\frac{1 - \cos \beta_0 h}{\sin \beta_0 h} \right)^2 = 120 \left(\tan \frac{\beta_0 h}{2} \right)^2. \quad (45)$$

APPENDIX III

With the notation (3), allowing γ to be complex and letting

$$z = e^{-\gamma s}, \quad (46)$$

the left-side member of (2) may be written

$$S(z) = Z_{00} + \sum_{n=1}^{\infty} Z_{0n} \left(z^n + \frac{1}{z^n} \right). \quad (47)$$

Consider now the series obtained from the series term of (47) upon replacing the mutual impedances by their asymptotic expressions (8). With the notation (9) we have

$$\begin{aligned}
S^{(a)}(z) &= S_1^{(a)}(z) + S_2^{(a)}(z) \\
&= jk \sum_{n=1}^{\infty} \frac{(x_0 z)^n}{n} + jk \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{x_0}{z} \right)^n. \quad (48)
\end{aligned}$$

$S_1^{(a)}(z)$ is convergent if

$$\text{either } |x_0 z| < 1 \text{ or } |x_0 z| = 1 \text{ and } x_0 z \neq 1. \quad (49)$$

$S_2^{(a)}(z)$ is convergent if

$$\text{either } \left| \frac{x_0}{z} \right| < 1 \text{ or } \left| \frac{x_0}{z} \right| = 1 \text{ and } \frac{x_0}{z} \neq 1. \quad (50)$$

Therefore, bearing in mind that $|x_0| = 1$, $S^{(a)}(z)$ is convergent if, and only if, the three following conditions are simultaneously satisfied,

$$|z| = 1, \quad z \neq x_0, \quad z \neq x_0^{-1}. \quad (51)$$

We will now prove that the same conditions hold for the series in (47).

Suppose that $S^{(a)}(z)$ is convergent; then by the comparison criterion it is sufficient to show that it is possible to fix two integers $N > 0$ and P such that, for any $n > N$, the following relationship is true:

$$|Z_{0n}| \leq \left| \frac{jk x_0^{n-P}}{n-P} \right| = \frac{k}{n-P}. \quad (52)$$

In fact it was shown in Appendix II that the difference

$$\left| Z_{0n} - \frac{jk x_0^n}{n} \right|$$

is of the order of $1/n^\alpha$ with $\alpha \geq 2$. We can therefore write, ϵ being a suitable constant,

$$|Z_{0n}| \leq \left| \frac{jk x_0^n}{n} \left(1 + \frac{\epsilon}{n^{\alpha-1}} \right) \right| \leq \frac{k}{n} \left(1 + \frac{|\epsilon|}{n^{\alpha-1}} \right). \quad (53)$$

The inequality (52) may then be replaced by the following:

$$\frac{1}{n} + \frac{|\epsilon|}{n^\alpha} \leq \frac{1}{n-P}. \quad (54)$$

From this one gets

$$P \geq \frac{|\epsilon|}{n^{\alpha-2} + \frac{|\epsilon|}{n}}. \quad (55)$$

This condition can be satisfied for any $n \geq N$ if $\alpha \geq 2$, by taking, for example,

$$P \geq \frac{|\epsilon|}{N^{\alpha-2}}. \quad (56)$$

In the same way it can be shown that two integers $N' > 0$ and P' exist such that, for any $n > N'$,

$$|Z_{0n}| \geq \left| \frac{jkx_0^{n+P'}}{n+P'} \right| = \frac{k}{n+P'}. \quad (57)$$

Thus, if (48) is divergent, (47) is also divergent and the proof is complete.

On physical grounds the preceding considerations are quite obvious; no attenuation can be present in a wave traveling along an infinite uniform array, if the wave is different from zero throughout the array, unless a source of infinite power is present, and in this case the assumptions that led to (2) no longer hold.

ACKNOWLEDGMENT

The authors wish to acknowledge the valuable experimental work contributed to this study by E. K. Damon.

Network Theory and Its Relation to the Theory of Linear Systems

J. MEIXNER†

Summary—Some general theorems of network theory are studied in their relation to other fields of linear physics. It is pointed out that general laws of the same character hold for the dissipative linear systems (thermodynamic systems with small departures from equilibrium). Another class of linear systems which one meets in scattering problems can also be related to network theory, but in a different way. For an example in which both the theory of dissipative and of scattering systems applies, the results of the two approaches are compared. Finally, results are mentioned on almost equivalent networks which have a bearing on the transition from the reversible behavior of the atomistic constituents of matter to the irreversible behavior of matter in bulk.

THE purpose of this paper is to enunciate some theorems on electrical networks and to demonstrate their relations to the general theory of linear systems (in particular to the theory of relaxation phenomena), to the scattering problems in classical and quantum physics, and to statistical mechanics.

Two kinds of linear systems are being considered in physics. Systems of the first kind are thermodynamic systems which undergo small changes such that they always remain close to a fixed equilibrium state. Phenomena observed in such systems, under this condition, are usually called relaxation phenomena. Networks have often been used as models for relaxation phenomena. Actually, networks are nothing else but thermodynamic systems of the kind mentioned; and, therefore, the general properties of networks also derive from a general theory of relaxation phenomena.

Scattering systems constitute systems of the second kind; examples of this are the reflection and diffraction of acoustic or electromagnetic waves and the scattering of elementary particles by nuclei, etc., in the language of wave mechanics. However, an electrical network can also be used to construct a scattering system of a special kind, if it is put at the terminals of a transmission line and if the reflection properties of a signal put on the line are studied.

The properties of an electrical network are determined by its structure. If it is contained in a "black box," such that only the terminals are accessible—we restrict the following to two-terminal networks, though the generalization to $2n$ -terminal networks is obvious and immediate—the properties are fully expressed by its impedance function $Z(p)$, where p is the time-differentiation operator, $p = d/dt$. If $u(t)$ is the applied voltage, and $i(t)$ the current through the terminals, we then have $u(t) = Z(p)i(t)$.

One can evaluate $Z(p)$ from the structure of the network. The converse is not true; for given $Z(p)$, there are many different structures with the same impedance. They are called equivalent networks. A famous example is given by the two networks in Fig. 1. From this example, one concludes that there is not even a correspondence in the sum of electrical and magnetic energies of equivalent networks. This means that the sum of these energies cannot be expressed in terms of the impedance function.

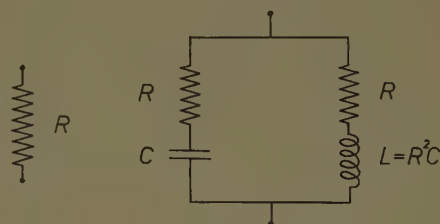


Fig. 1—Example for two networks with the same impedance function.

Tellegen¹ has, however, demonstrated that equivalent networks with the same input voltage $u(t)$ (where $u(t) = 0$ for $t < 0$), have at every moment the same difference of the electrical and the magnetic energy. From this, one expects that this difference may be expressible by the impedance function $Z(p)$. In fact, one derives for any input $u(t)$, and the resulting current $i(t)$, the identity

$$\frac{1}{2} \frac{Y(p') - Y(p)}{p' - p} u(t)u(t') = \sum \frac{1}{2C_k} q_k(t)q_k(t') - \frac{1}{2} \sum \sum L_{kl} \dot{i}_k(t)\dot{i}_l(t'), \quad (1)$$

where $p = d/dt$, $p' = d/dt'$ and $Y(p) = 1/Z(p)$. The C_k are the capacitances and the $q_k(t)$ their charges, the L_{kl} the inductances and the i_k, i_l the currents through the corresponding elements. In the limit $t' \rightarrow t$, the right member of (1) is the difference of the whole electric and magnetic energies stored in the network, while the left member contains only those quantities which pertain to the outside of the "black box."

Only in the case of reactance systems can one also make a statement on the total electric and magnetic energies in the systems in terms of the impedance function. This is trivial, however, since it is a direct consequence of the energy law.

† Institut für Theoretische Physik, Technische Hochschule Aachen, Aachen, Germany.

¹ B. D. H. Tellegen, "A general network theorem, with applications," *Philips Res. Repts.*, vol. 7, pp. 259-269; 1952.

There is a general theorem on the representation of the impedance of a network by an after-effect function $P(t)$. It also applies, under very loose restrictions, to networks with continuous elements, and it states that the impedance $Z(p)$ of any network can be expressed as

$$Z(p) = Ap + \frac{B}{p} - P'(+0) + \int_0^{\infty} e^{-pt} [P(t) - P''(t)] dt, \quad (2)$$

with $A \geq 0$, $B \geq 0$, while $P(t)$ is a real positive definite function. This representation permits the expression of the difference of the electric and magnetic energy of the network, as given in (1), by the after-effect function $P(t)$. The result is

$$W_{el} - W_{mag} = \frac{B}{2} q(t)^2 - \frac{A}{2} i(t)^2 + \frac{1}{2} \int_0^{\infty} d\tau_1 \int_0^{\infty} d\tau_2 i(t - \tau_1) i(t - \tau_2) \cdot [P(\tau_1 + \tau_2) - P''(\tau_1 + \tau_2)], \quad (3)$$

if $q(t)$ denotes the time integral of the current $i(t)$, with $q(-\infty) = 0$.²

The theory of linear networks is a special case of the theory of linear dissipative systems. Such systems are thermodynamic systems which always remain close to a fixed equilibrium state. If one knows the molecular structure of such a system, that is the molecular mechanisms which become active when one works from the outside on the systems by changing the intensive parameters (for instance pressures of pistons attached to the system, but also electric or magnetic fields), then the thermodynamically conjugated extensive parameters (displacements of the pistons, electric displacement, and magnetic induction) can be evaluated. For small departures from equilibrium, the relation between the changes of the intensive and extensive parameters are linear. In the following we restrict the discussion to one intensive and one extensive parameter and assume that both are zero in equilibrium.

If the molecular structure of the system is not known, one can, nevertheless, arrive at some conclusions concerning the relation between the intensive parameter $g(t)$ and the extensive parameter $f(t)$. We start from three evident postulates:

1) The superposition principle holds for the functions $f(t)$ and $g(t)$; that is, one is derived from the other by a linear transformation.

2) Translation invariance of time holds; that is, if $g(t)$ is the linear transform of $f(t)$, then $g(t+\tau)$ with any real τ is the linear transform of $f(t+\tau)$.

3) Causality is true; that is, if $f(t)$ is zero from $-\infty$ to some t , then $g(t)$ has the same property.

² J. Meixner, "Impedanz- und Lagrange-Funktion linearer dissipativer System," *Z. Physik*, vol. 156, pp. 200-210; 1952.

In fact, these postulates have been used for more than one hundred years to give a basis for the theory of linear system.³ As a matter of fact, they are, however, too general to specify the admitted linear transformations in dissipative linear systems. They apply to such systems as well as to scattering systems and, therefore, do not emphasize the characteristic properties of one or the other kind of systems.

In the case of dissipative systems, one must and can add a very powerful postulate, of a quite general nature, which makes even causality superfluous, as a separate postulate.⁴ Rather, causality is an immediate consequence of the new postulate. This is true, not only in the direction stated in postulate 3), but also in the reverse direction; that is, with $f(t)$ and $g(t)$ interchanged in postulate 3).

The basis of this postulate is the second law of thermodynamics, according to which the irreversible work done by changing the extensive parameters from the initial equilibrium state to some final values is always greater than the reversible work done in the same process. We express this by the postulate:⁵

4) The system is dissipative in the sense of thermodynamics, or

$$\int_{-\infty}^{\tau} g(t) \frac{df(t)}{dt} dt \Big|_{\text{irreversible}} \geq \int_0^{f(\tau)} g df \Big|_{\text{reversible}}. \quad (4)$$

This inequality expresses the fact that the irreversible work done on a system is greater than the work done in a reversible process, which leads to the same final value of the extensive variable f . In the second integral g is understood to be the equilibrium value $g(f)$ at given constant f .

In network theory, postulate 4) is very well known in a somewhat weaker form, namely

$$\int_{-\infty}^t u(t) i(t) dt \geq 0, \quad (5)$$

if the network was initially empty; that is if it did not contain electric or magnetic energy. It is also well known that (5) leads to the property that the impedance $Z(p)$ is a "positive function" (that is, it is analytic for $\text{Re } p > 0$, it is real for positive real p , and it has positive real part for $\text{Re } p > 0$). For thermodynamic systems with adiabatic isolation or with constant temperature, one can weaken postulate 4) to make it similar to (5) and obtain the stronger form (4) by deduction, so that this postulate becomes less a statement about dissipation

³ J. M. C. Duhamel, "Mémoire sur la méthode générale relative au mouvement de la chaleur dans les corps solides plongés dans les milieux dont la température varie avec le temps," *J. école Polytech.*, Paris, vol. 14, pp. 20-77; 1833; L. Boltzmann, "Zur Theorie der elastischen Nachwirkung," *Ann. Phys.*, vol. 7, pp. 624-654; 1876.

⁴ This was remarked upon for the first time by D. C. Youla, L. J. Castriota and H. J. Carlin, "Scattering Matrices and the Foundations of Linear, Passive Network Theory," *Microwave Res. Inst., Polytechnic Inst. of Brooklyn, Brooklyn, N. Y.*, Rept. No. R-594-57, PIB-522; 1957.

⁵ J. Meixner, "Thermodynamische Erweiterung der Nachwirkungstheorie," *Z. Physik*, vol. 139, pp. 30-143; 1954.

than about energy conservation. This has the further advantage that the existence of an equilibrium value $g(f)$ for constant need not be required.

In order to derive conclusions from these postulates, one has to specify the character of the admitted functions $f(t)$ and $g(t)$. To avoid complications introduced by the use of step functions $f(t)$, which may give rise to singular functions $g(t)$, one restricts the input functions $f(t)$ to the class of functions with an infinite number of continuous derivatives and which decrease so fast for $t \rightarrow -\infty$, that

$$t^n f(t) \rightarrow 0 \text{ for } t \rightarrow -\infty \text{ and for every positive } n.$$

There is now a very loose and physically plausible restriction on the response functions $g(t)$; they are assumed to be at least continuous; that is, one permits the loss of even an infinite number of derivatives in the linear transformation $f(t) \rightarrow g(t)$. It turns out, however⁶—and postulate 4) is essential to that—that at most three derivatives are lost; that is, one can extend the class of functions $f(t)$ to those possessing three continuous derivatives; then $g(t)$ is at least continuous. A further extension would now be possible, but would necessitate the introduction of singular functions (step functions and their derivatives).

No additional condition has to be imposed on the class of linear transformations $f(t) \rightarrow g(t)$. This class is well defined as a consequence of postulate 4)⁷ and characterized by the representation (8) of the impedance.

From these postulates one derives (see König and Meixner⁶), that for the considered systems there exists an impedance function $Z(p)$ such that

$$g(t) = pZ(p)f(t), \quad (6)$$

with p understood as the operator d/dt , or written in Laplace transforms

$$\int_0^\infty e^{-pt} g(t) dt = pZ(p) \int_0^\infty e^{-pt} f(t) dt, \quad (7)$$

if $f(t) = g(t) = 0$ for $t < 0$ and $\text{Re } p > 0$. The impedance function $Z(p)$ turns out to be a positive function with the general properties as mentioned above and it admits for $\text{Re } p > 0$ quite generally the representation

$$Z(p) = Ap + \frac{B}{p} + \int_0^\infty e^{-pt} P(t) dt + p^2 \int_0^\infty e^{-pt} [P(0) - P(t)] dt. \quad (8)$$

One has $A \geq 0$, $B \geq 0$ while $P(t)$, which has the character of an after-effect function, is a so-called positive definite function and can be represented by a Fourier-Stieltjes integral

⁶ H. König and L. Meixner, "Lineare systeme und lineare transformationen," *Math. Nachrichtentech.*, vol. 19, pp. 265–322; 1958; J. Meixner and H. König, "Zur theorie der linearen dissipativen systemen," *Rheolog. Acta*, vol. 1, pp. 190–193; 1958.

⁷ H. König, "Zur theorie der linearen dissipativen transformationen," *Arch. Mathematik*, in press.

$$P(t) = \int_{-\infty}^{+\infty} e^{i\rho t} d\phi(\rho) \quad \text{for all real } t \quad (9)$$

with a real distribution function $\phi(\rho)$, which is monotonically increasing and bounded, continuous at $\rho = 0$ and odd. If both A and B are zero, one has to assume $\phi(+\infty) - \phi(-\infty) > 0$ in order that $Z(p) \neq 0$. The sum of the two integrals in (8) is itself a positive function. The after-effect function $P(u)$ expresses how the value $g(t)$ is influenced by the values of f at former times. This is clear from the relation

$$g(t) = Af''(t) + Bf(t) + \int_0^\infty f'(t-u)P(u)du + \int_0^\infty f'''(t-u)[P(0) - P(u)]du \quad \text{for all } t,$$

which follows from (6) and (8). If, in particular, $P(u)$ has two continuous derivatives in $0 < u < \infty$ with finite limits for $u \rightarrow 0$ and $P''(u)$ is bounded from above by a polynomial in u in the same interval (as can be usually assumed for physical systems) two partial integrations can be carried through in the last integral to give

$$g(t) = Af''(t) + Bf(t) - P'(0)f'(t) + \int_0^\infty f'(t-u)[P(u) - P''(u)]du \quad \text{for all } t.$$

This is equivalent to the simplification (2) of (8).

If one identifies $g(t)$ with the voltage $u(t)$ applied to a network of impedance $Z(p)$ and $f'(t)$ with the current $i(t)$, and if, furthermore, one chooses for $i(t)$ the current impulse $\delta(t)g_0$, where $\delta(t)$ is a unit impulse at $t = 0$ and g is a constant of the dimension of a charge, one obtains

$$u(t) = [B + P(t) - P''(t)]q_0 \quad \text{for } t > 0,$$

while $u(t)$ at $t = 0$ is a superposition of the singular functions $[A\delta'(t) - P'(0)\delta(t)]q_0$. This shows the connection of the function $P(t)$ with the impulse response.

In special cases, simpler representations are possible. If, for instance, the system behaves like a capacitance for slow and for abrupt changes of $f(t)$, more precisely, if

$$0 < B = \lim_{p \rightarrow 0} pZ(p) \leq \lim_{p \rightarrow \infty} pZ(p) < \infty, \quad (10)$$

one obtains

$$Z(p) = \frac{B}{p} + \int_0^\infty e^{-pt} Q(t) dt, \quad (11)$$

where $Q(t) = P(t) - P''(t)$ is again a positive definite function.

Many general results are obtained in this way for material coefficients like elastic moduli, viscosity, dielectric constant (or dielectric tensor) and magnetic permeability (or permeability tensor), as functions of frequency. In particular for the dielectric constant ϵ and the magnetic permeability μ , it results that

$$p\epsilon(p) \quad \text{and} \quad p\mu(p)$$

are positive functions. The special statements for real frequencies ω follow from this by putting $p = i\omega$ (if the time dependence of the electric field is assumed as $\exp(i\omega t)$). This is equivalent to saying that the imaginary parts of $\epsilon(i\omega)$ and $\mu(i\omega)$ have the opposite signs from the frequency.

The possibility of portraying a material with a frequency dependent dielectric constant by an electric network, as is often done in the theory of relaxation phenomena, stems from the fact that $p\epsilon(p)$ can always be approximated by a rational positive function and such a function can always be realized by a finite electric network with capacitances, resistances and inductances.⁸

From the functions $p\epsilon(p)$ and $p\mu(p)$ one cannot draw separate conclusions about the electric and magnetic field energies during a change of the fields, but the difference of these energies can be analogously expressed as in the case of electrical networks. Only for dispersion without absorption can the individual energies be expressed in terms of the field strengths and $\epsilon(p)$ and $\mu(p)$.

The theory of linear scattering systems can be developed on a postulutory basis in a way similar to that used before, $f(t)$ and $g(t)$ now representing the amplitudes of the oncoming and the reflected, scattered, or transmitted wave. If the underlying laws for the wave propagation are linear with coefficients that do not depend on time, the first three postulates are guaranteed. It may be remarked, however, that contrary to dissipative systems, the converse of postulate 3) is not true: If the reflected wave is zero for $t < 0$, the oncoming wave may very well have values different from zero during some time interval anterior to $t = 0$.

Postulate 4) has to be replaced by a different postulate 4') of similar power. It can also be formulated as a conservation law, which is in some cases equivalent to the energy law, but can be more generally expressed as a conservation law for the number of particles.

In order to be more specific without undue digression, the discussion is restricted to the reflection of an electromagnetic wave, which is incident normally on a homogeneous dielectric halfspace, and it is further assumed that it has a time dependence $\exp(i\omega t)$. If $f(t, x_1)$ is the incident wave at a point x_1 , $g(t, x_2)$ the reflected wave at a point x_2 , then the postulate 4') says that the reflection coefficient of the amplitude, $S(i\omega) = g(t, x_1)/f(t, x_2)$, has an absolute value either less than or equal to one.

The theory developed on the basis of these postulates, together with some mathematical requirements, shows that the function $S(i\omega)$ can be viewed as the boundary values of an analytic function $S(p)$, defined in the half plane $\text{Re } p > 0$ with the property $|S(p)| < 1$.

From this, it follows that $[1 - S(p)]/[1 + S(p)]$ is a positive function as explained before.

The famous dispersion relations⁹ which recently have been found to have so many applications in the theory of elementary particles, relate the real and the imaginary parts of $S(i\omega)$ with each other. It must be remarked here that dispersion relations generally apply to scattering systems. They do not generally hold for dissipative systems, that is, for the real and imaginary parts of an impedance function. This is because an impedance function is not necessarily bounded in absolute value on the frequency axis. If it is bounded, however, it can be formally considered as a scattering function and the dispersion relations hold.

There are problems in which both the theory of dissipative systems and the theory of scattering systems can be applied. An example has been given in the previous section. The reflection coefficient of an electromagnetic wave incident normally on a dielectric halfspace with dielectric constant $\epsilon(p)$ and permeability $\mu(p)$ is, as is well known,

$$S = \frac{1 - \sqrt{\mu/\epsilon}}{1 + \sqrt{\mu/\epsilon}}. \quad (12)$$

From this it follows that

$$\sqrt{\frac{\mu}{\epsilon}} = \frac{1 - S}{1 + S}. \quad (13)$$

Therefore, $[\mu(p)/\epsilon(p)]^{1/2}$ has to be a positive function, as pointed out in the last section. This result is, however, contained in the result for dissipative systems, applied to the material of the dielectric halfspace, that $p\epsilon(p)$ and $p\mu(p)$ are positive functions. Even by considering more complicated situations, such as a dielectric layer of arbitrary thickness, eventually coated on one side with a perfectly conducting sheet, one does not obtain results which go beyond the one obtained from dissipativity.

Therefore, it seems as if the theory of dissipative linear systems has at least the same power as the theory of linear scattering systems, in cases where both theories are applicable. The first theory has, at least in this example, the advantage of a much easier access to the result.

In concluding, the author would like to examine an entirely different subject, which is also clarified by scrutinizing the properties of electrical networks. It concerns the old question: Why does matter exhibit irreversible behavior in the sense of the second law of thermodynamics, while the underlying laws governing the motion of the particles constituting the matter, are

⁸ R. Bott and R. J. Duffin, "Impedance Synthesis without Use of Transformers," *J. Appl. Phys.*, vol. 20, p. 816; August, 1949.

⁹ H. A. Kramers, "La diffusion de la lumière par les atomes," *Atti Congr. dei Fisici, Como 2*, pp. 545-557; 1927. R. Kronig, "On the theory of dispersion of x-rays," *J. Opt. Soc. Amer.*, vol. 12, pp. 547-557; November, 1926.

reversible? To this end we consider electrical networks which one might call almost equivalent networks.

By almost equivalent networks we mean networks which show for rather arbitrary inputs $u(t)$ (which are zero for $t < 0$) almost the same responses $i(t)$ in some fixed time interval $0 < t < t_1$. There is a general theorem on the representation of dissipative networks by almost equivalent networks, which do not contain resistances. It is a direct consequence of the expression (8) for the impedance function, with $P(t)$ inserted from (9), if one approximates the integral by a finite sum.

This discussion is restricted to giving a simple example of two almost equivalent networks, one of them being dissipative, while the other one is of the reactance type. Let us assume that the C_j and L_j in the second network of Fig. 2 are related to the C and R of the first network by

$$C_j = \frac{2C}{\pi} \frac{x}{1 + j^2 x^2}, \quad L_j C_j = \frac{\tau^2}{j^2 x^2} \quad (j = 1, 2, \dots, n). \quad (14)$$

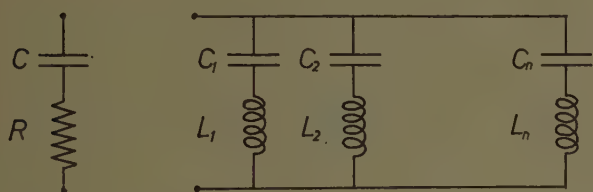


Fig. 2—Example for two almost equivalent networks.

While $\tau = RC$, the quantities x and n are available parameters. Let us further assume that the impressed voltage for both networks is $u(t)$ with $u(t) = 0$ for $t < 0$ and that the total charge transmitted after $t = 0$ is $q(t)$ for the left, $q_{n,x}(t)$ for the right network. Then one derives

$$|q(t) - q_{n,x}(t)| \leq C \left[\frac{t}{\tau} x + \frac{4}{n\pi x} \right] \int_0^t |u'(s)| ds \quad \text{for all } t \geq 0. \quad (15)$$

Since x can be made arbitrarily small and n arbitrarily large simultaneously, one has $q_{n,x}(t)$ approximately equal to $q(t)$ in an interval $0 < t < t_1(n, x, U)$, for all inputs $u(t)$ for which $\int_0^{t_1} |du/dt| dt$ is less than some value U . Apparently t_1 increases with n , $1/x$ and $1/U$.

From this example one can clearly see how reversibility and irreversibility can be reconciled. Irreversibility in reversible systems is bound by the requirement that the number of elements (here $2n$; in mechanical systems one would have to take the number of atomic constituents) must be very large and the observation time is restricted. Furthermore, one observes that the converse problem (under what conditions does a reversible system show irreversibility?) seems to be more or less a problem concerning the limiting behavior of the spectrum of eigenfrequencies of the reversible system, when its size increases indefinitely. There are simple examples of this, some of which will appear in a future publication which is being prepared by the author.

Linear Arrays: Currents, Impedances, and Fields, I*

RONOLD KING†

Summary—The radiation field of dipoles is usually expressed in the form of a product of a field factor and one or more array factors. Actually, such a formulation depends on the implicit assumption that the distributions of current along all elements are the same regardless of the location in the array or differences in driving conditions. Since this is a satisfactory approximation only when the elements are near a half-wavelength long, a study of the fields of arrays of longer elements in terms of the actual distributions of current is indicated.

The available solutions of the simultaneous integral equations for the distributions of current in a circular array of N -parallel elements are adequate in a quantitative sense only for arrays of approximately half-wave elements. Moreover, the form of the solution as a series of complicated terms is quite useless (except in the leading sinusoidal term) when the determination of the radiation field is desired. For this reason, a new solution for the currents, which provides a good approximation for the currents in arrays of elements that may be a full wavelength long or longer, and which is expressed in terms of combinations of simple trigonometric functions, is derived.

The new solution is used to determine the currents in and the fields of isolated half-wave and full-wave dipoles more accurately than with the conventional sinusoidal currents. The currents, impedances, and fields of a two-element array of full-wave elements are studied under various driving conditions including the broadside, bilateral end-fire, unilateral end-fire or couplet, and the case when one of the elements is parasitic. It is shown that the null in the conventional pattern for the couplet with identically distributed currents becomes a minor maximum with an amplitude equal to half that of the principal maximum when the more accurate distributions of current are used. The significance of this fact in its application to the minor lobe structure of more general arrays is considered. The usefulness of the new theory in determining the radiation field of parallel arrays, in general and when scanned, is discussed, and plans for further work are outlined.

THE CONVENTIONAL FORMULATION

THE far-zone electromagnetic field of an array of N structurally identical dipole antennas, each of half-length h , is generally expressed by the product of a field factor $F(\theta, \beta h)$ characteristic of a single element and an array factor $\alpha(\theta, \Phi)$ which is determined by the geometry of the array. Thus, the electric field is represented in the form,

$$E_\theta^r(\theta, \Phi) = KF(\theta, \beta h)\alpha(\theta, \Phi), \quad (1)$$

where K is a constant, θ and Φ are spherical coordinates, $\beta = 2\pi/\lambda$, and the superscript r denotes the radiation field. Frequently the array factor is itself the product of several partial array factors. For example, for an array which includes rows of broadside and end-fire elements, the array factor may have the form,

$$\alpha(\theta, \phi) = \alpha_B(\theta, \phi)\alpha_E(\theta, \Phi).$$

Specific formulas are found in King.¹ In so-called *uniform* arrays, the currents in the regularly arranged elements must be equal in amplitude, but they may differ uniformly and progressively in phase. More generally, in Tchebycheff, binomial, parasitic, and other *non-uniform* arrays, the currents in the several elements are not equal.

Whenever the field of an array is expressed as the product of a field factor and an array factor, it is implied that the individual elements *all* have the *same* field factor $F(\theta, \beta h)$ regardless of their locations in the array, the relative amplitudes and phases of the driving voltages, and the impedances of the generators or loads which are connected across the terminal pairs at the centers of the elements. In general, this tacit assumption is an approximation that varies in its accuracy with the degree in which differences in relative geometry and in conditions at the driving point affect the relative distributions of current along the elements. For most arrays constructed of half-wave dipoles, the field patterns calculated from the simple product formulas are adequate for many applications. For arrays with longer elements, such as full-wave dipoles, and for all arrays when the detailed minor-lobe structure of the pattern is quantitatively important, a more accurate representation of the radiation field than that provided by the conventional product formulas is required. Such a representation is, of course, already available, in a formal sense, in the explicit sum of N terms from which (1) is derived by assuming that $F(\theta, \beta h)$ is a common factor. Each of the terms in this sum gives the actual contribution to the field by the currents at all points in each element under operating conditions. Specifically, for an array of N -nonstaggered elements all parallel to the z axis of a system of Cartesian coordinates that has its origin at the center of the array (or at another convenient reference point), a general expression for the electric field at a point P in the radiation zone is

$$E_\theta^r(\theta, \Phi) = -j\omega A_\theta^r(\theta, \Phi) = -j\omega A_s^r(\theta, \Phi) \sin \theta, \quad (2)$$

where $A^r(\theta, \Phi) = \mathbf{A}_s^r(\theta, \Phi)$ is the resultant vector potential at P due to the actual currents in all of the elements in the array. The spherical coordinate θ is measured from the z axis and the coordinate Φ from the x axis. Specifically,

$$A_s^r(\theta, \Phi) = \sum_{i=1}^N A_{si}^r, \quad (3)$$

* The research reported was performed under Contract No. AF 19(604)-4118 between the AF Cambridge Research Center and the President and Fellows of Harvard College.

† Cruft Laboratory, Harvard University, Cambridge, Mass.

¹ R. W. P. King, "Theory of Linear Antennas," Harvard University Press, Cambridge, Mass., ch. 6, pp. 600-601; 1956.

$$A_{zi}r = \frac{\mu_0}{4\pi} \frac{e^{-i\beta R_i}}{R_i} \int_{-h}^h I_{zi}(z') e^{i\beta z' \cos \theta_i} dz'. \quad (4)$$

Instead of determining the resultant vector potential or electric field from the currents actually maintained in the array in normal operation, the principle of superposition may be used, as suggested by Hines, *et al.*³ This requires the separate and successive determination of the phase and amplitude of the entire field when each element is driven, one at a time, by its assigned voltage and the subsequent superposition of the N fields. (If the antennas are excited by current generators with assigned currents, the method of superposition involves the determination of the N fields obtained when each element is excited alone and in turn by the specified current while all others are left open-circuited at the terminals. Superposition again gives the desired pattern.) From the analytical point of view, this procedure is useful in calculating the far-zone field of an arbitrarily driven array only if the actual distributions of current in the single driven element, and in all of the $N-1$ parasitic elements, are known when each of the N elements is, in its turn, the driven one. Actually, Hines *et al.*³ proposed no method for determining these currents analytically or for calculating the field patterns. Their suggestion was to *measure*, in amplitude and phase, the N field patterns obtained by successively driving each element with the rest of the array present but parasitic, and then superimposing the patterns so obtained with appropriately assigned impressed input currents or driving voltages. While such a method is certainly superior to any cut-and-try procedure, it requires extensive experimental facilities. Moreover, it is analytically unsatisfactory and throws no light on the significance of the comments made by Hines *et al.*³ which suggest that the differences between (1) and (2) with (3) and (4) are not related to "the problem of mutual impedance between the elements, or the problem of holding currents in individual elements to assigned values in the presence of mutual impedance, or any problem concerning the net-

Diagram illustrating the geometry for calculating the path length difference S_i for an antenna in an array. The diagram shows a 3D coordinate system with axes x , y , and z . A point P is located at a distance R from the origin O . A vector R_i connects the origin O to a point on the surface of a sphere of radius r_i . The angle between the z -axis and the vector R is θ . The angle between the x -axis and the projection of R_i onto the xy -plane is ϕ_i . The path length difference S_i is the distance between the point on the sphere and the point P . The diagram also shows the vector A_z and the angle $-A_\theta$.

$$S_i = r_i \cos(\Phi - \Phi_i) \sin \theta$$
$$A_z^r(\theta, \Phi) = \frac{\mu_0}{4\pi} \frac{e^{-i\beta R}}{R} \sum_{i=1}^N e^{i\beta S_i} \int_{-h}^h I_{zi}(z') e^{i\beta z' \cos \theta} dz', \quad (5)$$
$$R_i = R - S_i, \quad S_i = r_i \sin \theta \cos (\Phi - \Phi_i), \quad (6)$$
$$I_{zi}(z') = I_{zi}(0)f_i(z') = I_{zm}(0)k_{mi}e^{-j\delta_m}f_i(z'), \quad (7)$$

where k_m and δ_m are the normalized real-amplitude and relative-phase angles of the current at the center of an-

³ J. N. Hines, V. H. Rumsey, and T. E. Tice, "On the design of arrays," *PROC. IRE*, vol. 42, pp. 1263-1265; August, 1954.

tenna i as referred to the current at the center of the reference antenna m . Also let

$$F_m(\theta, \beta h) = \int_{-h}^h f_m(z') e^{i\beta z' \cos \theta} \sin \theta dz', \quad (8)$$

$$G_{mi}(\theta, \beta h) = \int_{-h}^h [f_i(z') - f_m(z')] e^{i\beta z' \cos \theta} \sin \theta dz'. \quad (9)$$

Substitutions from (7), (8), and (9) in (5) lead to the following result:

$$A_z^r(\theta, \Phi) = K_m [F_m(\theta, \beta h) \alpha(\theta, \Phi) + D_{mi}(\theta, \Phi; \beta h)] \quad (10)$$

where

$$K_m = \frac{\mu_0 I_{zm}(0)}{2\pi \sin \theta} \frac{e^{-i\beta R}}{\beta R},$$

$$\alpha(\theta, \Phi) = \sum_{i=1}^N k_{mi} e^{i(\beta S_i - \delta_{mi})} \quad (11b)$$

$$D_m(\theta, \Phi; \beta h) = \sum_{i=1}^N k_{mi} e^{i(\beta S_i - \delta_{mi})} G_{mi}(\theta, \beta h). \quad (11c)$$

In the general formula (10) for the far-zone vector potential of an array of N -parallel dipoles with their centers in the plane $z=0$, the first term is the conventional representation in the usual product form and the second term is the correction. The integral $F_m(\theta, \beta h)$ is the field factor of antenna m in the presence of the array with a specified set of driving conditions. Note that if these conditions are changed, all of the $f_i(z')$ including $f_m(z')$ may be altered and, therefore, $F_m(\theta, \beta h)$ may also be altered. The sum $\alpha(\theta, \Phi)$ is the array factor of a parallel array of N elements of which all have identical field factors like that of element m .

In the special case of a broadside array with an odd number of elements all driven in such a manner that their input currents are equal and in phase, $k_{mi}=1$, $\delta_{mi}=0$ for all values of i . With the central unit as reference, and with $S_i = ib \sin \theta \cos \Phi$ as shown in Fig. 2(a), it follows that

$$\alpha_B(\theta, \Phi) = \frac{\sin (\frac{1}{2} N \beta b \sin \theta \cos \Phi)}{\sin (\frac{1}{2} \beta b \sin \theta \cos \Phi)} \quad (12)$$

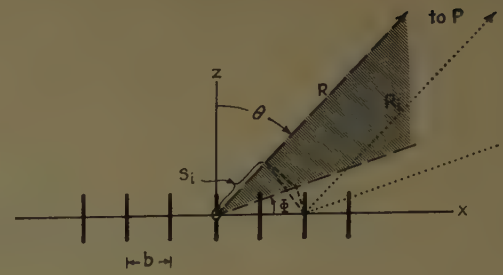
$$D_i(\theta, \Phi; \beta h) = \sum_{i=1}^{(N-1)/2} 2G_i(\theta, \beta h) \cos \beta S_i. \quad (13)$$

On the other hand, for an array consisting of an even number of elements all uniformly spaced around a circle of radius c , as shown in Fig. 2(b), and with driving-point currents which are equal and in phase, the array factor and correction term are given by

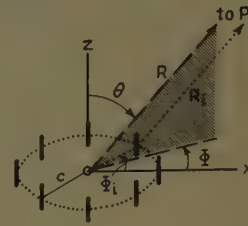
$$\alpha(\theta, \Phi) = 2 \sum_{i=1}^{N/2} \cos \{ \beta c \sin \theta \cos [\Phi - 2\pi(i-1)/N] \} \quad (14a)$$

$$= J_0(\beta c \sin \theta) \quad \text{for } N > 6 \quad (14b)$$

$$D_{1i}(\theta, \Phi; \beta h) = 0. \quad (15)$$



(a)



(b)

Fig. 2—(a) Broadside array, $N=7$; (b) circular array, $N=8$.

Note that the correction term vanishes identically since $f_i(z') = f_m(z')$ for all values of i and choices of m . This follows from the obvious fact that the N units are geometrically and electrically identical, so that not only the input currents, but the entire distributions of current on all antennas, must be the same. Thus, for the uniformly driven circular array, the simple product of a field factor times an array factor is *exact* within the approximations of a far-zone formula. Note that this is true for a circular array with an odd number of units, and also for the case when the elements are driven by voltages which are equal in magnitude but which increase progressively, and uniformly, in phase from element to element around the circle. However, $F_m(\theta, \beta h)$ is, in general, not the same for arrays having different values of the phase shift from element to element.

It may be concluded that the simple product of a field factor and an array factor differs from the correct representation of the far-zone field of an array by a term $D_{mi}(\theta, \Phi; h)$ which depends on the *differences* between the normalized distributions of current along the several elements. Such differences must exist in greater or lesser degree whenever the electrical environments of the individual elements are not identical. Note, however, that the effect of dissimilar environments in modifying the distributions of current also depends on the natural condition of resonance as determined by the common length of the elements.

INTEGRAL EQUATIONS FOR THE CURRENTS

The determination of the distributions of current along the N -geometrically identical antennas of a parallel array depends upon the solution of N -simultaneous integral equations which are derived by imposing the boundary condition $E_z = 0$ on the cylindrical surfaces of the several antennas. The details are found in King.⁴

⁴ King, *op. cit.*, p. 351 ff.

The final step is the equating of two different expressions for the vector potential $A_z(z)$ on the surface of each antenna. Thus, for antenna k ,

$$4\pi\nu_0 A_{zk}(z) = \int_{-h}^h \sum_{i=1}^N I_{zi}(z') K_{ki}(z, z') dz' \\ = \frac{-j4\pi}{\zeta_0} [C_k \cos \beta z + \frac{1}{2} V_{0k} \sin \beta |z|], \quad (16)$$

where

$$K_{ki}(z, z') = \frac{e^{-j\beta R_{ki}}}{R_{ki}}, \quad R_{ki} = \sqrt{(z_k - z_i')^2 + b_{ki}^2}, \quad b_{kk} = a, \\ \nu_0 = \frac{1}{\mu_0} = \frac{10^7}{4\pi} \text{ meters/henry}, \quad \zeta_0 \doteq 120\pi \text{ ohms.} \quad (17)$$

R_{ki} is the distance from a point at z on the surface of antenna k , where $A_{zk}(z)$ is calculated to a point at z' on the axis of antenna i where the element of integration dz' is located. The distance between the centers of antennas k and i is b_{ki} ; the radius of each antenna is a . V_{0k} is the driving voltage⁵ at the center, $z=0$, of antenna k ; C_k is a constant of integration that must be evaluated from the condition $I_{zk}(h)=0$. As a consequence of symmetry, $I_{zk}(-z) = I_{zk}(z)$. The several distances and the notation are illustrated in Fig. 3 for an array of three antennas.

If the N voltages, V_{0k} , $k=1, 2, \dots, N$, are assigned, (16) represents N simultaneous integral equations for the N currents, $I_{zk}(z) = I_{zk}(0)f_k(z)$. Alternatively, the N driving-point currents $I_{zk}(0)$, $k=1, 2, \dots, N$, may be assigned while the N driving voltages, V_{0k} , and the N distribution functions, $f_k(z)$, are the unknowns. In general, the N simultaneous integral equations (16) have not been solved. In the special case of the circular array, and with the help of the method of symmetrical components, they may be replaced by an equivalent set of N independent integral equations for N phase-sequence currents.⁶ The N phase-sequences are obtained by driving the N antennas with voltages which are related as follows in the adjacent elements k and $k+1$: $V_{k+1} = V_k e^{(j2\pi m/N)}$, where m may have the values $m=0, 1, 2, \dots, N-1$. Note that in the zero phase-sequence ($m=0$), all voltages are in phase. The vector potential and integral equation for the m th phase-sequence current $I^{(m)}(z)$ in any of the N elements when each is driven by the voltage $V_0^{(m)}$, are

$$4\pi\nu_0 A_z^{(m)}(z) = \int_{-h}^h I^{(m)}(z') K^{(m)}(z, z') dz' \\ = \frac{-j4\pi}{\zeta_0} [C^{(m)} \cos \beta z + \frac{1}{2} V_0^{(m)} \sin \beta |z|], \quad (18)$$

⁵ A δ -function generator is assumed here. The singularity in susceptance inherent in such a generator is, in effect, subtracted out in the approximate solution by iteration. A detailed analysis of this problem is found in T. T. Wu and R. W. P. King, "Driving point and input admittance of linear antennas," *J. Appl. Phys.*, vol. 30, pp. 74-76; January, 1959.

⁶ King, *op. cit.*, ch. 3, sec. 14.

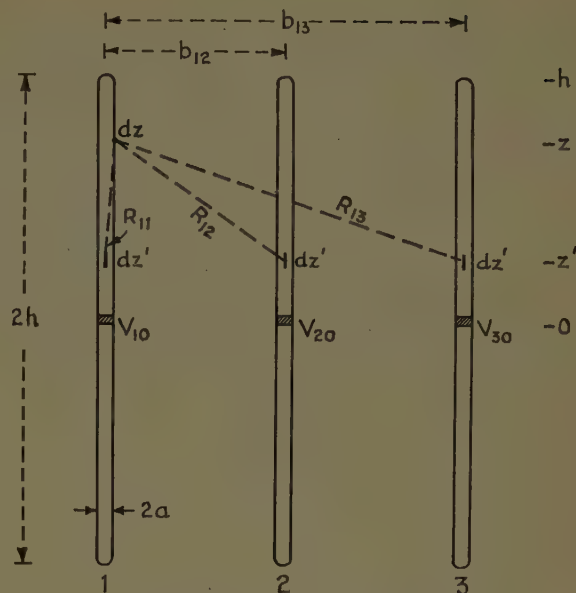


Fig. 3—Three-element array.

where

$$K^{(m)}(z, z') = \sum_{i=1}^N p^{(i-1)m} \frac{e^{-j\beta R_{1i}}}{R_{1i}} \quad (19)$$

with

$$p = e^{j2\pi/N}, \quad R_{1i} = \sqrt{(z_1 - z_i')^2 + b_{1i}^2}, \quad b_{11} = a. \quad (20)$$

The equations for the several phase sequences are obtained successively from (20) by assigning the values $m=0, 1, 2, \dots, N-1$. By a suitable superposition of the phase-sequence currents obtained by solving (18) for all N values of m , the current distribution in each element of a circular array may be obtained when the elements are driven by arbitrary voltages. The driving-point impedances are then readily obtained from the currents in the input terminals.

It is not possible to obtain a set of independent integral equations like (18) for noncircular parallel arrays. To be sure, the current $I_{zk}(z')$ can be factored out in the integrand of (16) and the remaining terms can be combined into a kernel in order to achieve formal equivalence with (18). However, in this case, the kernel is not independent of the distributions of current in the several elements so that nothing is gained by the rearrangement. Note, however, that the kernel becomes a known function if the currents $I_{zi}(0)$, $i=1, 2, \dots, N$, at the driving points of all elements are specified and if it is assumed as an approximation that the distribution of current along each coupled element is like that along antenna k so that for all values of i , $I_{zi}(z) = I_{zk}(z)$. In this manner (16) is converted into an independent integral equation for determining the current distribution $I_{zk}(z)$ and the driving voltage V_{0k} . By allowing k to assume all values from 1 to N , independent integral equations are obtained for all N -current distributions and driving voltages. Unlike (18), however, these equations are all approximate since it has been assumed, in each, that the currents in all

coupled antennas are distributed exactly as in the element in question. This is the same approximation as that implied in the product form of the far-zone field of the array. It is satisfactory only in certain special cases.

The effect of coupled antennas on the distributions of currents in the elements of an array may be studied in the circular array for which a fundamentally correct formulation is available, provided a tractable solution can be found. Since the integral equation (18) for the current $I^{(m)}(z)$ in the m th-phase sequence of the circular array is exactly like the integral equation for the current in an isolated cylindrical antenna except for a different kernel, it might be supposed that the same method of iteration⁷ should be applicable. Actually, this is not true if $K^{(m)}(z, z')$ as defined in (19) is merely substituted for $K_1(z, z') = e^{-j\beta R_{11}}/R_{11}$ in the various integrals of the solution for the current in the isolated antenna. Such a substitution would violate the fundamental principle of the method; that the ratio $A_z(z)/I_z(z)$ of the vector potential on the surface to the total current in the antenna is sensibly constant and predominantly real. Except in the case of very closely spaced elements, the vector potential along antenna k due to currents in antenna i does not vary, in general, with z as does $I_{iz}(z)$. Hence, unless the coupled antennas are very few in number, or so close together that they constitute a cage or a multiwire transmission line, or so far apart that mutual contributions to the vector potentials are very small, the ratio of vector potential to current cannot be expected to be even moderately constant along any of the elements.

Fortunately, this difficulty may be side-stepped in an alternative arrangement of the integral equation (18) in which the vector potential $A_z^{(m)}(z)$ on the surface of each element is separated into two parts, of which one is constant along the entire antenna while the other is variable but vanishes at the ends. The desired form is readily obtained from (18) by introducing the vector potential difference, $W_z^{(m)}(z) = A_z^{(m)}(z) - A_z^{(m)}(h)$, together with the constant,

$$U^{(m)} = \frac{-j\omega}{\beta} A_z^{(m)}(h) = \frac{-j\zeta_0}{4\pi} \int_{-h}^h I_z^{(m)}(z') K^{(m)}(h, z') dz'. \quad (21)$$

The desired equivalent form is

$$4\pi\nu_0 W_z^{(m)}(z) = \int_{-h}^h I_z^{(m)}(z') K_d(z, z') dz' = \frac{-j4\pi}{\zeta_0} [C^{(m)} \cos \beta z + \frac{1}{2} V_0^{(m)} \sin \beta |z| + U^{(m)}], \quad (22)$$

where the new difference kernel is

$$K_d^{(m)}(z, z') = K^{(m)}(z, z') - K^{(m)}(h, z') = \sum_{i=1}^N p^{(i-1)m} \left[\frac{e^{-j\beta R_{1i}}}{R_{1i}} - \frac{e^{-j\beta R_{1ih}}}{R_{1ih}} \right], \quad (23)$$

⁷ *Ibid.*, ch. 2, sec. 14.

and where

$$R_{1ih} = \sqrt{(h - z_i')^2 + b_{1i}^2}, \quad b_{11} = a. \quad (24)$$

Note that at $z=h$, (22) reduces to

$$C^{(m)} \cos \beta h + \frac{1}{2} V_0^{(m)} \sin \beta h + U^{(m)} = 0. \quad (25)$$

This relation may be used to eliminate $U^{(m)}$, as in the conventional formulation for a single antenna, or it may be used to determine $C^{(m)}$. In the first case (22) assumes Form I:

$$\int_{-h}^h I_z^{(m)}(z') K_d^{(m)}(z, z') dz' = \frac{-j4\pi}{\zeta_0} [C^{(m)} F_{0z} + \frac{1}{2} V_0^{(m)} G_{0z}], \quad (26)$$

where

$$F_{0z} = F_0(z) - F_0(h) = \cos \beta z - \cos \beta h \quad (27)$$

$$G_{0z} = G_0(z) - G_0(h) = \sin \beta |z| - \sin \beta h. \quad (28)$$

Alternatively, if (25) is used to eliminate $C^{(m)}$, (22) assumes Form II:

$$\int_{-h}^h I_z^{(m)}(z') K_d^{(m)}(z, z') dz' = \frac{j4\pi}{\zeta_0 F_0(h)} [U^{(m)} F_{0z} + \frac{1}{2} V_0^{(m)} M_{0z}], \quad (29)$$

where

$$M_{0z} = F_0(z) G_0(h) - G_0(z) F_0(h) = \sin \beta (h - |z|). \quad (30)$$

Each of these equations can be solved by iteration. If enough terms are used, they must yield essentially the same result. The problem at hand, however, is not primarily that of obtaining an accurate solution expressed in terms of complicated integrals that require numerical methods of evaluation but, rather, that of deriving a reasonably accurate expression for the distribution of current which takes proper account of the contributions of coupled antennas and which is useful in evaluating the radiation field. It will be shown that the two equivalent integral equations, (26) and (29), lead to quite different trigonometric approximations of the currents.

SOLUTION OF FORM I OF THE INTEGRAL EQUATION

The rearranged integral equation (26) is preferred to the original equation (18) because the ratio $W_z^{(m)}(z)/I_z^{(m)}(z)$ should be more nearly constant as a function of z in the range $-h \leq z \leq h$ than the ratio $A_z^{(m)}(z)/I_z^{(m)}(z)$. This is a consequence of the fact that $W_z^{(m)}(z)$ includes only the part of $A_z^{(m)}(z)$ which varies in this range and not the part which is constant. Moreover, large variations in $W_z^{(m)}(z)$ from contributions by the current in a coupled element can occur only when this element is quite near, in which case $W_z^{(m)}(z)$ varies approximately as the adjacent current maintaining it. This reasoning suggests that the method of iteration used to solve the equation for the isolated antenna

should be applicable to the formally identical equation (26) of the circular array driven in its m th phase sequence. The procedure used is the introduction of an approximate (and, as yet, undetermined) distribution function $g(z, z') \doteq I_s^{(m)}(z')/I_s^{(m)}(z)$ in the integral for the vector potential difference and so define the difference function,

$$\Psi_d(z) = \Psi(z) - \Psi(h) = \int_{-h}^h g(z, z') K_d^{(m)}(z, z') dz' = \Psi_d + \gamma(z), \quad (31)$$

where Ψ_d is the desired constant and $\gamma(z)$ is a small correction term. It is then argued that the difference,

$$D(z) = \Psi_d I_s^{(m)}(z) - \int_{-h}^h I_s^{(m)}(z') K_d^{(m)}(z, z') dz', \quad (32)$$

must be small. It follows directly that by adding $\Psi_d I_s^{(m)}(z)$ to both sides of (26) and by moving the integral to the right, a form suitable for iteration is obtained. Thus,

$$I_s^{(m)}(z) = \frac{-j4\pi}{\zeta_0 \Psi_d} \left[C^{(m)} F_{0z} + \frac{1}{2} V_0^{(m)} G_{0z} \right] + \frac{D(z)}{\Psi_d}. \quad (33)$$

A suitable zero-order term is that given in brackets in (33). The successive substitution of this term for the current in the small difference $D(z)$ and the evaluation of $C^{(m)}$ by a similar substitution in $U^{(m)}$ in (25), gives

$$I_s^{(m)}(z) = \frac{j2\pi V_0^{(m)}}{\zeta_0 \Psi_d} \cdot \left[\frac{M_{0z} + M_{1z}/\Psi_d + M_{2z}/\Psi_d^2 + \dots}{F_0(h) + F_1(h)/\Psi_d + F_2(h)/\Psi_d^2 + \dots} \right]. \quad (34)$$

The first-order functions in (34) are

$$M_{1z} = F_{1z} G_0(h) - G_{1z} F_0(h), \quad (35)$$

$$F_{1z} = F_{0z} \Psi_d - \int_{-h}^h F_{0z'} K_d^{(m)}(z, z') dz' \quad (36a)$$

$$= F_{0z} \Psi_d - [C_z(h, z) - C_z(h, h)] + [E_z(h, z) - E_z(h, h)] F_0(h), \quad (36b)$$

$$F_1(h) = - \int_{-h}^h F_{0z'} K^{(m)}(h, z') dz' \quad (37a)$$

$$= -C_z(h, h) + E_z(h, h) F_0(h). \quad (37b)$$

The functions G_{1z} and $G_1(h)$ are, respectively, like (36a), (36b), (37a) and (37b) with F replaced by G , and C replaced by S . The following function has been defined:

$$C_z(h, z) = C_a(h, z) + \sum_{i=2}^N p^{(i-1)m} C_{bi}(h, z). \quad (38)$$

The function $S_z(h, z)$ or $E_z(h, z)$ is obtained from (38) by replacing C by S or C by E . The three functions [in which $R_1 = \sqrt{(z-z')^2 + a^2}$, $R_2 = \sqrt{(z+z')^2 + a^2}$],

$$C_a(h, z) = \int_{-h}^h \cos \beta z' \left[\frac{e^{-j\beta R_1}}{R_1} + \frac{e^{-j\beta R_2}}{R_2} \right] dz' \quad (39a)$$

$$S_a(h, z) = \int_{-h}^h \sin \beta z' \left[\frac{e^{-j\beta R_1}}{R_1} + \frac{e^{-j\beta R_2}}{R_2} \right] dz' \quad (39b)$$

$$E_a(h, z) = \int_{-h}^h \left[\frac{e^{-j\beta R_1}}{R_1} + \frac{e^{-j\beta R_2}}{R_2} \right] dz', \quad (39c)$$

may be expressed entirely in terms of tabulated generalized sine- and cosine-integral functions. Explicit formulas are found in King.⁸ The kernels $K_d^{(m)}(z, z')$ and $K^{(m)}(h, z')$ are defined in (23). Second-order integrals are obtained from (36a) and (37a) by changing the subscripts zero to 1, and 1 to 2.

The zero-order term in (34) should be a suitable approximate distribution of current for use in the evaluation of Ψ_d in (31). With $g(z, z') = M_{0z'}/M_{0z}$ it follows that

$$\Psi_d(z) = M_{0z}^{-1} \{ [C_z(h, z) - C_z(h, h)] G_0(h) - [S_z(h, z) - S_z(h, h)] F_0(h) \}. \quad (40)$$

The success of the method of iteration now depends on whether or not the function $\Psi_d(z)$ is essentially constant in the range $-h \leq z \leq h$ at a value Ψ_d . It is known that, for an isolated antenna,

$$\Psi_d M_{0z} \doteq [C_a(h, z) - C_a(h, h)] G_0(h) - [S_a(h, z) - S_a(h, h)] F_0(h), \quad (41)$$

is a good approximation when Ψ_d is defined by

$$\Psi_d = \begin{cases} |\Psi_d(0)| & \text{for } \beta h \leq \pi/2 \\ |\Psi_d(h - \lambda/4)| & \text{for } \beta h \geq \pi/2. \end{cases} \quad (42)$$

Investigation shows that, in general, (41) is also a reasonable approximation if the subscript a is replaced by Σ and if the definition of Ψ_d by (42) is retained. However, the approximation is not equally good under all conditions. It is very good when 1) the contributions by the terms in the sums are small because the number of elements is not large, or because all elements are sufficiently far apart, or both; 2) all the coupled elements are so close together in terms of the wavelength that they may be regarded as forming a cage antenna, a multi-wire transmission line, or a superposition of these; and 3) the elements are approximately half-wave dipoles. However, if no restrictions are placed on numbers or separations as in 1), the definition of Ψ_d may be improved over that stated in (42) by setting,

$$\Psi_d = \Psi_d(0) \quad \text{for } \beta h \sim \pi/2. \quad (43a)$$

Specifically, at $\beta h = \pi/2$,

$$\Psi_d = \left[C_z\left(\frac{\lambda}{4}, 0\right) - C_z\left(\frac{\lambda}{4}, \frac{\lambda}{4}\right) \right]. \quad (43b)$$

Note that (43a) allows Ψ_d to be complex. This is possible when $\beta h \sim \pi/2$ since the real and imaginary parts of

the vector potential difference have distributions that are similar to each other and to the current, and since both have their maxima near $z=0$.

In general, the approximation $\Psi_d(z) \doteq \Psi_d M_{0z}$ is least satisfactory when βh is near π , primarily because the component of current in phase quadrature with the driving voltage is distributed very differently from both the component of current in phase with the voltage and the induced currents, whatever their phases relative to $V_0^{(m)}$. Note that in (34) the leading term is in phase quadrature with the voltage if Ψ_d is real, as required by (42).

SOLUTION OF FORM II OF THE INTEGRAL EQUATION

The alternative Form II of the integral equation for the current in the m th phase sequence of a circular array, as given in (29), may be solved by a process of iteration similar to that already described but with certain refinements. These are suggested by the fact that the right-hand side of (29) is expressed formally in terms of two source functions $V^{(m)}$ and $U^{(m)}$, the one a potential difference localized at $z=0$ and the other a field of constant amplitude distributed over the entire length of each element. Clearly, the current $I_z^{(m)}(z)$ in each element may be regarded as the sum of two components, of which one is generated directly by $V^{(m)}$ much as if the antenna were isolated and the other is induced by $U^{(m)}$ as in a receiving antenna in a uniform field. The leading term in the part of the current maintained directly by $V^{(m)}$ is $M_{0z} = \sin \beta(h - |z|)$; the leading term in the current maintained by $U^{(m)}$ is $F_{0z} = \cos \beta z - \cos \beta h$. If the substitution $I_z^{(m)}(z) = I_V^{(m)}(z) + I_U^{(m)}(z)$ is made in the integral in (29), the terms may be arranged as follows:

$$\begin{aligned} & \int_{-h}^h I_V^{(m)}(z') K_d^{(m)}(z, z') dz' + \int_{-h}^h I_U^{(m)}(z') K_d^{(m)}(z, z') dz' \\ &= \frac{j4\pi}{\xi_0 F_0(h)} [U^{(m)} F_{0z} + \frac{1}{2} V_0^{(m)} M_{0z}]. \quad (44a) \end{aligned}$$

The integrals on the left in (44a) are proportional, respectively, to the vector potential differences $W_V^{(m)}(z)$ and $W_U^{(m)}(z)$ at the point z on antenna 1 due to the components of the current $I_V^{(m)}(z)$ and $I_U^{(m)}(z)$ in all of the elements both near and far. Contributions to these components of the vector potential by the currents in antenna 1 and in any other elements so near that $\beta b_{1i} < 1$, vary with z approximately like the currents that maintain them. On the other hand, contributions from all elements which are not very near to antenna 1 ($\beta b_{n+1} \geq 1$) vary with z in a manner determined by the over-all effect of the currents and not by their local distributions. This is illustrated in Fig. 4 in which the quantity $S_b(\lambda/2, z) - S_b(\lambda/2, \lambda/2)$ as defined in (39b) is shown. It is proportional to the vector potential difference $W_V^{(m)}(z)$ along an axis at a distance b from an antenna of half-length $h = \lambda/2$ with a current distributed according to

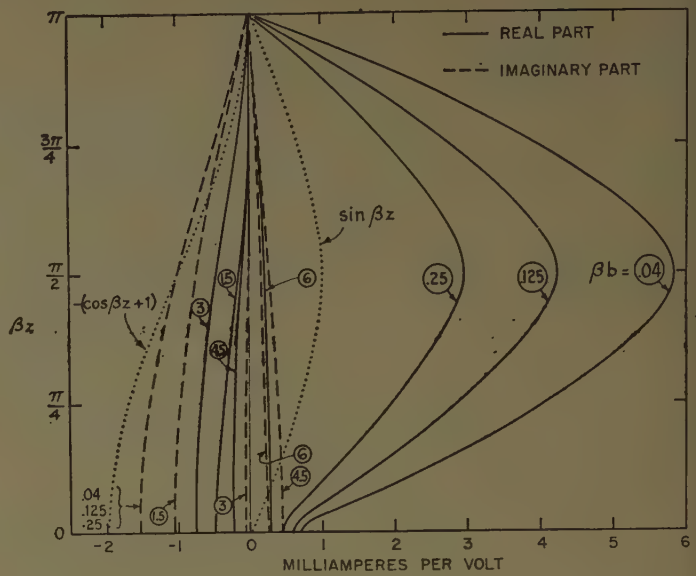


Fig. 4—The function $S_b(h, z) - S_b(h, h)$ for $h = \lambda/2$.

$\sin \beta z$. Note that, except near $z=0$, the real part varies very much like $\sin \beta z$ provided $\beta b < 1$. On the other hand, both the imaginary part for all values of βb , and the real part for $\beta b \geq 1$, vary approximately like $1 + \cos \beta z$. In a similar manner, it can be shown that the distribution of the vector potential $W_U^{(m)}(z)$ due to a current distribution $\cos \beta z + 1$, varies approximately like $\cos \beta z + 1$ for all values of βb in both real and imaginary parts. These facts suggest the following rearrangement of the integrals on the left in (44a):

$$\begin{aligned} & \int_{-h}^h I_V^{(m)}(z') K_{dR}^{(m)}(z, z') dz' + j \int_{-h}^h I_V^{(m)}(z') K_{dI}^{(m)}(z, z') dz' \\ &+ \int_{-h}^h I_V^{(m)}(z') K_{d\Sigma}^{(m)}(z, z') dz' + \int_{-h}^h I_U^{(m)}(z, z') K_d^{(m)}(z, z') dz' \\ &= \frac{j4\pi}{\xi_0 F_0(h)} [U^{(m)} F_{0z} + \frac{1}{2} V_0^{(m)} M_{0z}] \quad (44b) \end{aligned}$$

where $K_{dR}^{(m)}(z, z')$ and $K_{dI}^{(m)}(z, z')$ are, respectively, the real and imaginary parts of that part of $K_d^{(m)}(z, z')$, as defined in (23), for which $\beta b_{1i} < 1$, and where $K_{d\Sigma}^{(m)}(z, z')$ is the remaining part of $K_d^{(m)}(z, z')$; that is,

$$\begin{aligned} & K_{dR}^{(m)}(z, z') + j K_{dI}^{(m)}(z, z') \\ &= \sum_{i=1}^n p^{(i-1)m} \left[\frac{e^{-j\beta R_{1i}}}{R_{1i}} - \frac{e^{-j\beta R_{1ih}}}{R_{1ih}} \right]; \quad \beta b_n < 1; \quad (45a) \end{aligned}$$

$$\begin{aligned} & K_{d\Sigma}^{(m)}(z, z') \\ &= \sum_{i=n+1}^N p^{(i-1)m} \left[\frac{e^{-j\beta R_{1i}}}{R_{1i}} - \frac{e^{-j\beta R_{1ih}}}{R_{1ih}} \right]; \quad \beta b_{n+1} \geq 1. \quad (45b) \end{aligned}$$

It is now convenient to define three approximate distribution functions for the current as follows.

$$\begin{aligned} g_{VU}(z, z') &\doteq I_V^{(m)}(z') / I_U^{(m)}(z) T(h), \\ g_U(z, z') &\doteq I_U^{(m)}(z') / I_U^{(m)}(z), \\ g_V(z, z') &\doteq I_V^{(m)}(z') / I_V^{(m)}(z). \end{aligned}$$

The dimensionless constant $T(h)$ is introduced here for later convenience. Also, let the following functions be defined:

$$\Psi_{dR}(z) = \Psi_R(z) - \Psi_R(h) \\ = \int_{-h}^h g_V(z, z') K_{dR}(z, z') dz' = \Psi_{dR} + \gamma_R(z), \quad (46a)$$

$$\Psi_{dU}(z) = \Psi_U(z) - \Psi_U(h) \\ = \int_{-h}^h g_U(z, z') K_{dU}(z, z') dz' = \Psi_{dU} + \gamma_U(z), \quad (46b)$$

$$\Psi_{dI}(z) = \Psi_I(z) - \Psi_I(h) \\ = \int_{-h}^h g_V(z, z') K_{dI}(z, z') dz' = \Psi_{dI} + \gamma_I(z), \quad (46c)$$

$$\Psi_{dZ}(z) = \Psi_Z(z) - \Psi_Z(h) \\ = \int_{-h}^h g_U(z, z') K_{dZ}(z, z') dz' = \Psi_{dZ} + \gamma_Z(z). \quad (46d)$$

It is anticipated that, by a proper choice of $g(z, z')$ in each case, the integral may be made sensibly constant with the value Ψ_{dR} , Ψ_{dU} , Ψ_{dI} , or Ψ_{dZ} , while the functions $\gamma(z)$ are small. In this case the difference functions,

$$D_R(z) = \Psi_{dR} I_V^{(m)}(z) - \int_{-h}^h I_V^{(m)}(z') K_{dR}(z, z') dz', \quad (47a)$$

$$D_U(z) = \Psi_{dU} I_U^{(m)}(z) - \int_{-h}^h I_U^{(m)}(z') K_{dU}(z, z') dz', \quad (47b)$$

$$D_I(z) = \Psi_{dI} I_U^{(m)}(z)/T(h) - \int_{-h}^h I_V^{(m)}(z') K_{dI}(z, z') dz', \quad (47c)$$

$$D_Z(z) = \Psi_{dZ} I_U^{(m)}(z)/T(h) - \int_{-h}^h I_V^{(m)}(z') K_{dZ}(z, z') dz' \quad (47d)$$

are small. With (47a), (47b), (47c) and (47d), the integral equation (44b) may be rearranged as follows:

$$I_V^{(m)}(z) \Psi_{dR} + I_U(z)^{(m)} (\Psi_{dZ} + j \Psi_{dI})/T(h) + I_U^{(m)}(z) \Psi_{dU} \\ = \frac{j4\pi}{\xi_0 F_0(h)} [U^{(m)} F_{0z} + \frac{1}{2} V_0^{(m)} M_{0z}] \\ + D_R(z) + D_Z(z) + j D_I(z) + D_U(z). \quad (48)$$

The first term on the left in (48) varies with z approximately as does M_{0z} ; all other terms vary approximately as F_{0z} . This suggests the possibility of satisfying (48) in two parts as follows:

$$I_V^{(m)}(z) \Psi_{dR} = \frac{j2\pi V_0^{(m)}}{\xi_0 F_0(h)} M_{0z} + D_R(z), \quad (49a)$$

$$I_U^{(m)}(z) \Psi_{UI} = \frac{j4\pi U^{(m)}}{\xi_0 F_0(h)} F_{0z} + D_U(z) + D_Z(z) + j D_I(z), \quad (49b)$$

where

$$\Psi_{UI} = \Psi_{dU} + [\Psi_{dZ} + j \Psi_{dI}]/T(h). \quad (49c)$$

It is now possible to iterate (49a) for $I_V^{(m)}(z)$ and (49b) for $I_U^{(m)}(z)$. In the latter, use may be made of $I_V^{(m)}(z)$ in those higher-order integrals which depend upon it. The results are:

$$I_V^{(m)}(z) = \frac{j2\pi V_0^{(m)}}{\xi_0 \Psi_{dR} F_0(h)} m(z), \quad (50a)$$

$$I_U^{(m)}(z) = \frac{j4\pi U^{(m)}}{\xi_0 \Psi_{UI} F_0(h)} f(z), \quad (50b)$$

where

$$m(z) = M_{0z} + M_{R1z}/\Psi_{dR} + \dots, \quad (51a)$$

$$f(z) = F_{0z}[F_{1z} + H_{1z}/T(h)]/\Psi_{UI} + \dots \quad (51b)$$

In (51a), M_{R1z} is the real part of M_{1z} , as defined in (35), with Ψ_{dR} substituted for Ψ_d . In (51b), F_{1z} is given by (36) with Ψ_{dU} substituted for Ψ_d . The new function H_{1z} is defined as follows:

$$H_{1z} = F_{0z}(\Psi_{dZ} + j \Psi_{dI}) \\ - \int_{-h}^h M_{0z'} [K_{dZ}(z, z') + j K_{dI}(z, z')] dz'. \quad (51c)$$

The total current is the sum of (50a) and (50b), that is,

$$I_z^{(m)}(z) = \frac{j2\pi V_0^{(m)}}{\xi_0 \Psi_{dR} F_0(h)} [m(z) + T(h)f(z)], \quad (52)$$

where

$$T(h) = \frac{2U^{(m)} \Psi_{dR}}{V^{(m)} \Psi_{UI}}. \quad (53)$$

The constant $T(h)$, which relates the distributed field $U^{(m)}$ on the surface of each element to the driving voltages $V_0^{(m)}$ at their centers, may be determined readily by the substitution of (52) in the defining formula (21) for $U^{(m)}$. Thus,

$$U^{(m)} = \frac{V_0^{(m)} \Psi_{UI} T(h)}{2 \Psi_{dR}} \\ = \frac{V_0^{(m)}}{2 \Psi_{dR} F_0(h)} [\Psi_V(h) + T(h) \Psi_U(h)], \quad (54)$$

where

$$\Psi_V(h) = \int_{-h}^h m(z') K^{(m)}(h, z') dz', \quad (55a)$$

$$\Psi_U(h) = \int_{-h}^h f(z') K^{(m)}(h, z') dz'. \quad (55b)$$

The kernel $K^{(m)}(h, z')$ is defined as part of (23). The solution of (54) is

$$T(h) = \frac{\Psi_V(h)}{\Psi_{UV}F_0(h) - \Psi_U(h)} = \frac{\Psi_V(h)}{\{\Psi_{dU} + [\Psi_{dZ} + j\Psi_{dI}/T(h)]\}F_0(h) - \Psi_U(h)} \cdot (56)$$

The second expression is obtained with (49c). This equation may be solved for $T(h)$ to obtain

$$T(h) = \frac{\Psi_V(h) - [\Psi_{dZ} + j\Psi_{dI}]F_0(h)}{\Psi_{dU}F_0(h) - \Psi_U(h)}. (57)$$

The several Ψ -functions may now be evaluated in a first-order approximation by the substitution of the zero-order currents for the distribution functions. Thus,

$$g_{UV}(z, z') = M_{0z'}/F_{0z}; \quad g_V(z, z') = M_{0z'}/M_{0z}; \\ g_U(z, z') = F_{0z'}/F_{0z}. (58)$$

It follows that $\Psi_{dR}(z)$ is like the real part of $\Psi_d(z)$ as given in (40); hence,

$$\Psi_{dR} = \operatorname{Re} \begin{cases} \Psi_{dR}(0) & \text{for } \beta h \leq \pi/2 \\ \Psi_{dR}(h - \lambda/4) & \text{for } \beta h \geq \pi/2. \end{cases} (59a)$$

However, it is necessary to substitute C_{z1} for C_z and S_{z1} for S_z . Corresponding to (38), $C_{z1}(h, z)$ is defined by

$$C_{z1}(h, z) = C_a(h, z) + \sum_{i=2}^n p^{(i-1)m} C_{bi}(h, z); \quad \beta b_{1n} < 1. (59b)$$

Similarly, from (46b),

$$\Psi_{dU}(z) = F_{0z}^{-1} \{ [C_z(h, z) - C_z(h, h)] - [E_z(h, z) - E_z(h, h)]F_0(h) \}. (60a)$$

The appropriate choice of Ψ_{dU} for $\beta h \leq 2\pi$ is,

$$\Psi_{dU} = \Psi_{dU}(0) = \left\{ \frac{[C_z(h, 0) - C_z(h, h)] - [E_z(h, 0) - E_z(h, h)]F_0(h)}{1 - F_0(h)} \right\}. (60b)$$

From (46c) it follows that

$$\Psi_{dI}(z) = \operatorname{Im} F_{0z}^{-1} \{ [C_{z1}(h, z) - C_{z1}(h, h)]G_0(h) - [S_{z1}(h, z) - S_{z1}(h, h)]F_0(h) \}. (61a)$$

The appropriate choice of Ψ_{dI} is

$$\Psi_{dI} = \Psi_{dI}(0) = \operatorname{Im} \left\{ \frac{[C_{z1}(h, 0) - C_{z1}(h, h)]G_0(h) - [S_{z1}(h, 0) - S_{z1}(h, h)]F_0(h)}{1 - F_0(h)} \right\}. (61b)$$

It follows from (46d) that the parameter Ψ_{dZ} may be defined by

$$\Psi_{dZ} = \Psi_{dZ}(0) = \left\{ \frac{[C_{z2}(h, 0) - C_{z2}(h, h)]G_0(h) - [S_{z2}(h, 0) - S_{z2}(h, h)]F_0(h)}{1 - F_0(h)} \right\}, (62)$$

where, corresponding to (38),

$$C_{z2}(h, z) = \sum_{i=n+1}^N p^{(i-1)m} C_{bi}(h, z); \quad \beta b_{i,n+1} \geq 1. (63)$$

Note that Ψ_{dR} and Ψ_{dI} are real, whereas Ψ_{dU} and Ψ_{dZ} are complex.

The parameters $\Psi_V(h)$ and $\Psi_U(h)$ may be evaluated from (55a) and (55b) in a first-order approximation by the substitution of M_{0z} for $m(z)$ and F_{0z} for $f(z)$. The results are

$$\Psi_V(h) = C_Z(h, h)G_0(h) - S_Z(h, h)F_0(h) \quad (64)$$

$$\Psi_U(h) = C_Z(h, h) - E_Z(h, h)F_0(h). \quad (65)$$

These values, and the corresponding one from (61) and (63), may be substituted in (57) to obtain a first-order formula for $T(h)$. Important specific values are

$$T(\lambda/4) = -1, \quad (66a)$$

$$T(\lambda/2) = - \left\{ \frac{S_Z\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) + \frac{1}{2} \left[S_{Z2}\left(\frac{\lambda}{2}, 0\right) - S_{Z2}\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right] + j \operatorname{Im} \frac{1}{2} \left[S_{Z1}\left(\frac{\lambda}{2}, 0\right) - S_{Z1}\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right]}{\Psi_{dU} + C_Z\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) + E_Z\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right)} \right\}. \quad (66b)$$

The corresponding values of Ψ_{dU} and Ψ_{dR} are

$$\Psi_{dU} = C_Z\left(\frac{\lambda}{4}, 0\right) - C_Z\left(\frac{\lambda}{4}, \frac{\lambda}{4}\right) \text{ for } \beta h = \pi/2, \quad (67a)$$

$$\Psi_{dU} = \frac{1}{2} \left[C_Z\left(\frac{\lambda}{2}, 0\right) - C_Z\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) + E_Z\left(\frac{\lambda}{2}, 0\right) - E_Z\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right] \text{ for } \beta h = \pi, \quad (67b)$$

$$\Psi_{dR} = \operatorname{Re} \left[C_{Z1}\left(\frac{\lambda}{4}, 0\right) - C_{Z1}\left(\frac{\lambda}{4}, \frac{\lambda}{4}\right) \right] = \operatorname{Re} \Psi_{dU} \text{ for } \beta h = \pi/2, \quad (68a)$$

$$\Psi_{dR} = \operatorname{Re} \left[S_{Z1}\left(\frac{\lambda}{2}, \frac{\lambda}{4}\right) - S_{Z1}\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right] \text{ for } \beta h = \pi. \quad (68b)$$

Similarly,

$$\Psi_{dI} = \operatorname{Im} \left[C_{Z1}\left(\frac{\lambda}{4}, 0\right) - C_{Z1}\left(\frac{\lambda}{4}, \frac{\lambda}{4}\right) \right] \text{ for } \beta h = \pi/2, \quad (69a)$$

$$\Psi_{dI} = \operatorname{Im} \frac{1}{2} \left[S_{Z1}\left(\frac{\lambda}{2}, 0\right) - S_{Z1}\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right] \text{ for } \beta h = \pi, \quad (69b)$$

$$\Psi_{dZ} = \left[C_{Z2}\left(\frac{\lambda}{4}, 0\right) - C_{Z2}\left(\frac{\lambda}{4}, \frac{\lambda}{4}\right) \right] \text{ for } \beta h = \pi/2, \quad (70a)$$

$$\Psi_{dZ} = \frac{1}{2} \left[S_{Z2}\left(\frac{\lambda}{2}, 0\right) - S_{Z2}\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right] \text{ for } \beta h = \pi. \quad (70b)$$

The final solution for Form II of the integral equation is,

$$I_z^{(m)}(z) = \frac{j2\pi V_0^{(m)}}{\zeta_0 \Psi_{dR} F_0(h)} \left\{ [M_{0z} + T(h)F_{0z}] + \left[\frac{M_{R1z}}{\Psi_{dR}} + T(h) \frac{F_{1z} + H_{1z}/T(h)}{\Psi_{dU} + [\Psi_{dZ} + j\Psi_{dI}]/T(h)} \right] + \dots \right\}. \quad (71)$$

Because of the factor $F_0(h) = \cos \beta h$ in the denominator, this formula appears to have no application when $\beta h = \pi/2$. Actually, (71) does not become infinite but merely indeterminate of the form 0/0 at $\beta h = \pi/2$. It may be evaluated in the usual manner by differentiating numerator and denominator separately with respect to

βh and by setting $\beta h = \pi/2$. With (66a), the resulting formula is

$$I_z^{(m)}(z) = \frac{-j2\pi V_0^{(m)}}{\zeta_0 \Psi_{dR}} \left\{ \left[\sin \beta |z| - 1 + T'\left(\frac{\lambda}{4}\right) \cos \beta z \right] + \frac{d}{d\beta h} []_{\beta h = \pi/2} + \dots \right\}, \quad (72)$$

where $T'(\lambda/4)$ is the derivative with respect to βh of $T(h)$ evaluated at $\beta h = \pi/2$, and where the final square bracket is the same as that in (71). The differentiation of (57) gives

$$T'\left(\frac{\lambda}{4}\right) = \left[\Psi_{dU} - \Psi_{dZ} - j\Psi_{dI} - S_Z\left(\frac{\lambda}{4}, \frac{\lambda}{4}\right) + E_Z\left(\frac{\lambda}{4}, \frac{\lambda}{4}\right) \right] / C_Z\left(\frac{\lambda}{4}, \frac{\lambda}{4}\right) \quad (73)$$

where Ψ_{dU} , Ψ_{dI} , and Ψ_{dZ} are given by (67a) (69a), and (70a).

The formulas (71) and (72) for the current in the m th phase sequence of a circular array are preferred to the alternative form (34) primarily because, in them the leading trigonometric terms, with their complex coefficients, are constructed to provide a good approximation of both components of the current, whereas in (34) the leading trigonometric term yields only a poor approximation of one component of the current. These leading terms may be expanded as follows:

Form I:

$$[I_z^{(m)}(z)]_0 = \frac{j2\pi V_0^{(m)}}{\zeta_0 \Psi_d} \left(\frac{\sin \beta(h - |z|)}{\cos \beta h} \right). \quad (74)$$

Form II:

$$[I_z^{(m)}(z)]_0 = \frac{j2\pi V_0^{(m)}}{\zeta_0 \Psi_{dR}} \cdot \left(\frac{\sin \beta(h - |z|) + T(h)[\cos \beta z - \cos \beta h]}{\cos \beta h} \right), \quad \beta h \neq \pi/2 \quad (75a)$$

$$[I_z^{(m)}(z)]_0 = \frac{-j2\pi V_0^{(m)}}{\zeta_0 \Psi_{dR}} \left[T' \left(\frac{\lambda}{4} \right) \cos \beta z + \sin \beta |z| - 1 \right], \quad \beta h = \pi/2. \quad (75b)$$

THE ISOLATED LINEAR RADIATOR

In order to determine the accuracy of the new quasi-zero-order formulas (75a) and (75b) and in order to study their usefulness in calculating the radiation field, it is advantageous to apply them first to the simplest case; that is, the isolated antenna. For this, accurate theoretical and experimental admittances and distributions of current are available. A comparison of the far-zone fields calculated from the new formulas with those determined from the conventionally assumed, sinusoidally distributed current (74), is readily made. More accurate field patterns obtained from the complex effective length by application of the reciprocal theorem are also available.⁹

Calculations are conveniently made for a center-driven antenna for which $\Omega = 2\ln(2h/a) = 10$ with $\beta h = \pi/2$ and $\beta h = \pi$. The following data apply when $\beta h = \pi/2$:

$$\Psi_{dU} = C_a \left(\frac{\lambda}{4}, 0 \right) - C_a \left(\frac{\lambda}{4}, \frac{\lambda}{4} \right) = 7.688 - j0.633; \quad (76a)$$

$$\Psi_{dR} = \operatorname{Re} \left[C_a \left(\frac{\lambda}{4}, 0 \right) - C_a \left(\frac{\lambda}{4}, \frac{\lambda}{4} \right) \right] = 7.688; \quad (76b)$$

$$\Psi_{dI} = \operatorname{Im} \left[C_a \left(\frac{\lambda}{4}, 0 \right) - C_a \left(\frac{\lambda}{4}, \frac{\lambda}{4} \right) \right] = -0.633; \quad (76c)$$

$$\Psi_{dZ} = 0, \quad (76d)$$

$$T' \left(\frac{\lambda}{4} \right) = \left[\frac{\Psi_{dU} + j\Psi_{dI} - S_a \left(\frac{\lambda}{4}, \frac{\lambda}{4} \right) + E_a \left(\frac{\lambda}{4}, \frac{\lambda}{4} \right)}{C_a \left(\frac{\lambda}{4}, \frac{\lambda}{4} \right)} \right] = 3.153 + j4.553. \quad (77)$$

The substitution of these values in (75b) gives the following formula for the quasi-zero order current in amperes if V_0 is in volts.

$$[I_z(z)]_0 = V_0 \{ 9.87 \cos \beta z - j[6.84 \cos \beta z + 2.17(\sin \beta z - 1)] \} \times 10^{-3}. \quad (78)$$

The input admittance is

$$[Y_0]_0 = (9.87 - j4.67) \times 10^{-3} \text{ mhos}. \quad (79a)$$

The second-order theoretical value is

$$[Y_0]_2 = (9.26 - j4.62) \times 10^{-3} \text{ mhos}. \quad (79b)$$

(Note that the zero-order admittance obtained from (74) is $Y_0 = j\infty$.) The distribution of current as given by (78) is shown in Fig. 5; it is in excellent agreement with the approximate second-order curve shown in King.¹⁰

The following data apply when $\beta h = \pi$:

$$\Psi_{dU} = \frac{1}{2} \left[C_a \left(\frac{\lambda}{2}, 0 \right) - C_a \left(\frac{\lambda}{2}, \frac{\lambda}{2} \right) + E_a \left(\frac{\lambda}{2}, 0 \right) - E_a \left(\frac{\lambda}{2}, \frac{\lambda}{2} \right) \right] = 7.600 - j2.230, \quad (80a)$$

$$\Psi_{dR} = \operatorname{Re} \frac{1}{2} \left[S_a \left(\frac{\lambda}{2}, \frac{\lambda}{4} \right) - S_a \left(\frac{\lambda}{2}, \frac{\lambda}{2} \right) \right] = \operatorname{Re} (5.834 - j1.061) = 5.834, \quad (80b)$$

$$\Psi_{dI} = \operatorname{Im} \frac{1}{2} \left[S_a \left(\frac{\lambda}{2}, 0 \right) - S_a \left(\frac{\lambda}{2}, \frac{\lambda}{2} \right) \right] = \operatorname{Im} (0.353 - j0.779) = -0.779, \quad (80c)$$

$$\Psi_{dZ} = 0, \quad (80d)$$

and

$$T \left(\frac{\lambda}{2} \right) = - \left\{ \frac{S_a \left(\frac{\lambda}{2}, \frac{\lambda}{2} \right) + j \operatorname{Im} \frac{1}{2} \left[S_a \left(\frac{\lambda}{2}, 0 \right) - S_a \left(\frac{\lambda}{2}, \frac{\lambda}{2} \right) \right]}{\Psi_{dU} + C_a \left(\frac{\lambda}{2}, \frac{\lambda}{2} \right) + E_a \left(\frac{\lambda}{2}, \frac{\lambda}{2} \right)} \right\} = -0.169 + j0.174. \quad (81)$$

The current is obtained when these quantities are substituted in (75a). It is expressed in amperes in the following formula when V_0 is in volts:

$$[I_z(z)]_0 = V_0 \{ 0.497(\cos \beta z + 1) - j[2.805 \sin \beta z - 0.483(\cos \beta z + 1)] \} \times 10^{-3}. \quad (82)$$

⁹ King, *op. cit.*, pp. 488-493.

¹⁰ *Ibid.*, p. 116, Fig. 22.9.

The input admittance is

$$[Y_0]_0 = (0.994 + j0.966) \times 10^{-3} \text{ mhos.} \quad (83a)$$

The second-order theoretical value is

$$[Y_0]_2 = (1.00 + j1.80) \times 10^{-3} \text{ mhos.} \quad (83b)$$

[Note that the zero-order admittance obtained from (74) is $Y_0=0$.] The distribution of current as given by (82) is shown in Fig. 6; it is in very good agreement with the approximate second-order curve shown in King.¹¹

The general shapes of the curves of both components of the current shown in Figs. 5 and 6 for center-driven half-wave and full-wave dipoles are in remarkable agreement with higher-order theory and experiment if the fact that only the simplest trigonometric functions are involved is recalled. For $\beta h = \pi/2$, the quasi-zero-order input admittance as given in (77a) is within a few per cent of the second-order value given in (77b); when

¹¹ *Ibid.*, p. 117, Fig. 22.11.

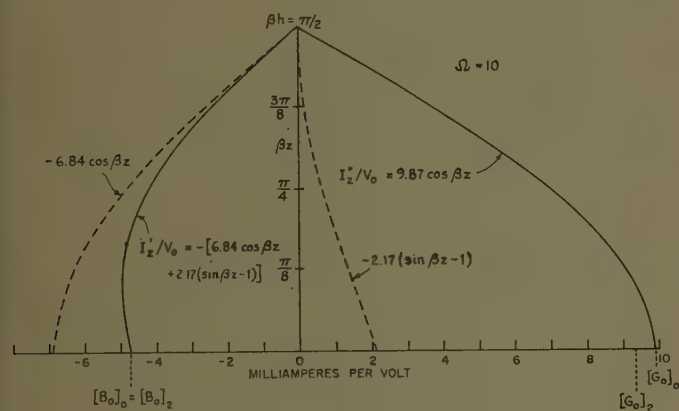


Fig. 5—Quasi-zero-order current.

$\beta h = \pi$, the agreement between the conductances in (81) and (82) is almost perfect, but the susceptances differ by more than 40 per cent. The reasons for this relatively large difference between zero- and second-order values of the susceptance when $\beta h = \pi$, are readily seen from a study of Fig. 7 in which the zero-order and first-order approximations of the first, second, and fourth integrals in (44b) are shown. From these it is clear that, for the shifted cosine distributions, the normalized amplitudes occur at $z=0$ whereas, for $\sin \beta z$, the same occurs at $z=h-\lambda/4$. As a result, the former, but not the latter, are most accurate at $z=0$. It would, of course, be possible to define an amplitude corresponding to Ψ_{dR} in terms of $\text{Re} [S_a(\lambda/2, z) - S_a(\lambda/2, 0)]$, instead of $\text{Re} [S_a(\lambda/2, z) - S_a(\lambda/2, \lambda/2)]$, in order to obtain an exact agreement of the zero- and first-order terms at $z=0$. However, the over-all agreement between the distribution curves would be poorer and the current would have a nonzero value at $z=h$. From the point of view of determining far-zone field patterns, an accurate

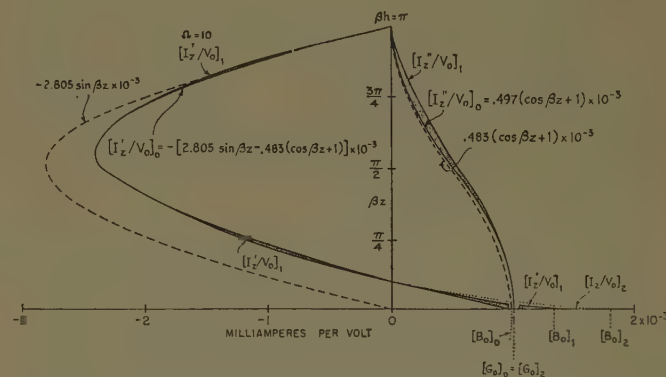


Fig. 6—Quasi-zero-order and higher-order currents (shaded areas show the difference between zero- and first-order values).

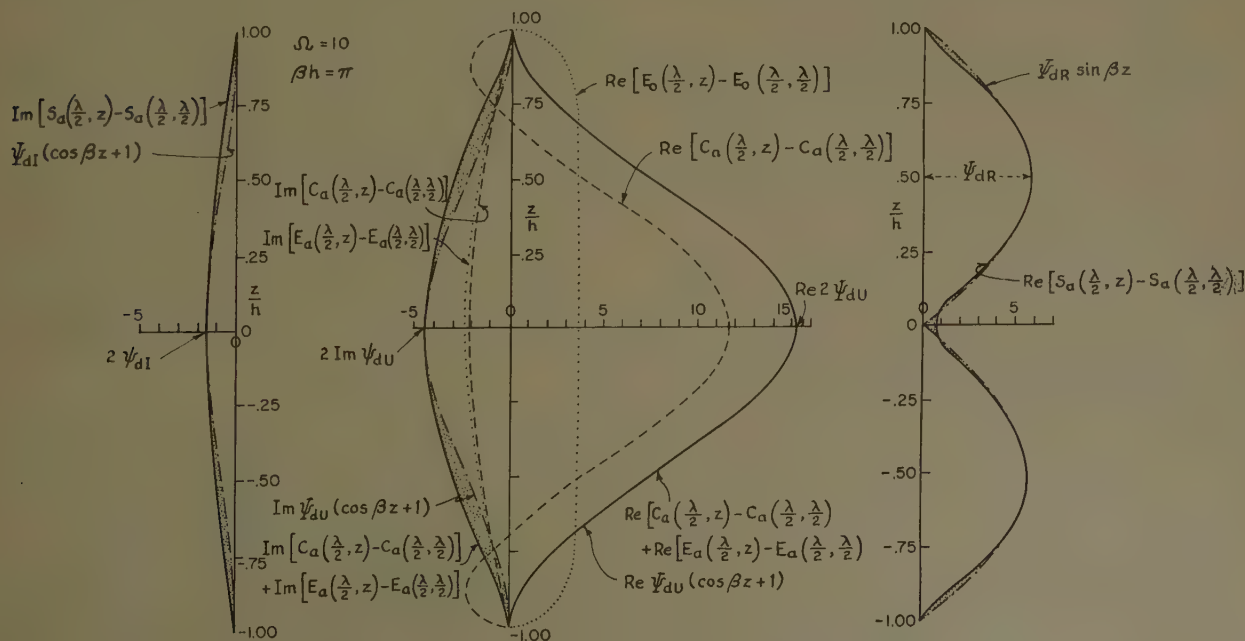


Fig. 7—First-order corrections of current (shown shaded).

representation of the currents along the antenna is of primary importance. For this purpose, the small change in current distribution, represented by a change in the input value, for agreement with the correct second-order input susceptance instead of the approximate zero-order value, is of no consequence. This is evident from Fig. 6 where the changes in current due to the first-order corrections are indicated by the shaded areas, and the additional change due to a second-order correction of the input susceptance is estimated. It is clear that the contributions to the field by these differences may be ignored. Moreover, since the input power is given to a high degree of accuracy by $[V_0^2 G_0]_0$, the quasi-zero-order current distribution may be used as an excellent approximation for all purposes except the specific determination of B_0 when βh is near π . In practice, the susceptance is, in any case, often greatly modified by terminal-zone effects. It is well to note that the quasi-zero-order representation of the current is least accurate at $z=0$ when $\beta h=\pi$ since the contribution by $\sin \beta(h-|z|)$ is then zero. For values of βh not near π , the real part of the function in the brace in (40) is much better approximated by $\Psi_{dR} \sin \beta(h-|z|)$ at $z=0$ than when $\beta h=\pi$.

The far-zone electric field of a linear radiator of length $2h$ is given by

$$E_{\theta}^r = -j\omega A_z^r \sin \theta \\ = -\frac{j\omega \sin \theta}{4\pi\nu_0} \frac{e^{-j\beta R_0}}{R_0} \int_{-h}^h I_z(z') e^{j\beta z' \cos \theta} dz'. \quad (84)$$

If the current distributions (75a) and (75b) are successively substituted in the integral in (84), the electric field may be expressed as follows:

$$E_{\theta}^r = \frac{V_0 e^{-j\beta R_0}}{\Psi_{dR} R_0 \cos \beta h} [F_m(\theta, \beta h) + T(h) G_m(\theta, \beta h)], \quad h \neq \pi/2; \quad (85a)$$

$$E_{\theta}^r = -\frac{V_0 e^{-j\beta R_0}}{\Psi_{dR} R_0} \left[T' \left(\frac{\lambda}{4} \right) F_m \left(\theta, \frac{\pi}{2} \right) + H_m \left(\theta, \frac{\pi}{2} \right) \right], \quad \beta h = \pi/2, \quad (85b)$$

where

$$F_m(\theta, \beta h) = \frac{\cos(\beta h \cos \theta) - \cos \beta h}{\sin \theta}, \quad (86a)$$

$$G_m(\theta, \beta h) = \frac{\sin \beta h \cos(\beta h \cos \theta) \cos \theta - \cos \beta h \sin(\beta h \cos \theta)}{\sin \theta \cos \theta}, \quad (86b)$$

$$H_m \left(\theta, \frac{\pi}{2} \right) = \frac{\cos \theta - \sin \left(\frac{\pi}{2} \cos \theta \right)}{\sin \theta \cos \theta}. \quad (86c)$$

Graphs of the functions $F_m(\theta, \pi)$ and $G_m(\theta, \pi)$ are shown in Fig. 8 together with the components of the electric field and its magnitude for a full-wave antenna with $\Omega = 2 \ln(2h/a) = 10$. Note that the instantaneous field

for a driving voltage $V_{inst} = V_0 \cos \omega t$ is

$$E_{\theta \cdot inst}^r = \frac{V_0}{\Psi_{dR} R_0} [-F_m(\theta, \pi) \cos(\omega t - \beta R_0) + G_m(\theta, \pi) \sin(\omega t - \beta R_0)]. \quad (87)$$

It is significant that, with the same unit amplitude of current, the value $F_m(\pi/2, \pi) = 2$ whereas $G_m(\pi/2, \pi) = \pi$.

Graphs of the field functions $F_m(\theta, \pi/2)$ and $H_m(\theta, \pi/2)$, which occur in (85b) for the far-zone field of the half-wave dipole, are shown in Fig. 9. It is seen that they differ so little in shape that the approximation

$$H_m \left(\theta, \frac{\pi}{2} \right) \doteq \left(\frac{\pi}{2} - 1 \right) F_m \left(\theta, \frac{\pi}{2} \right) \quad (88)$$

is excellent. With (88), (84b) reduces to the simple conventional form,

$$E_{\theta}^r \doteq A F_m \left(\theta, \frac{\pi}{2} \right), \quad (89)$$

where

$$A = \left[T' \left(\frac{\lambda}{4} \right) + \frac{\pi}{2} - 1 \right] \frac{V_0}{\Psi_{dR}} \frac{e^{-j\beta R_0}}{R_0}. \quad (90)$$

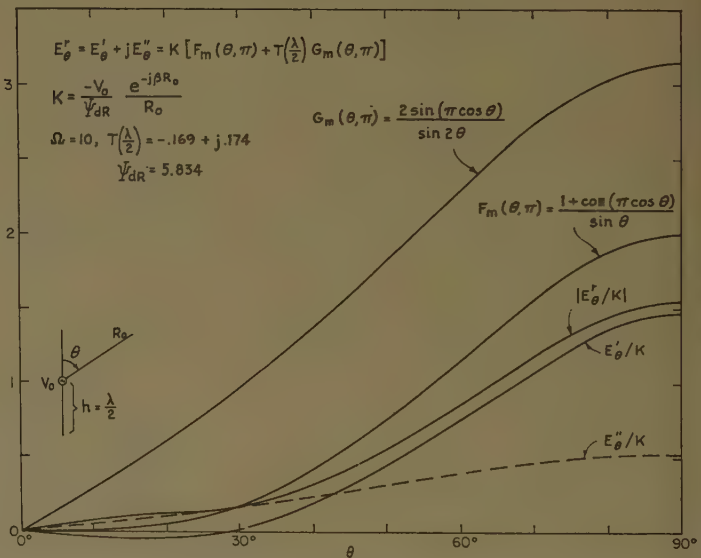


Fig. 8—Vertical field pattern of full-wave center-driven antenna.

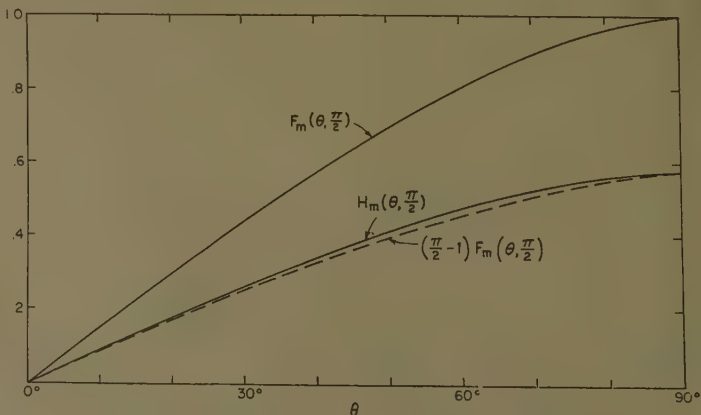


Fig. 9—The field functions $F_m(\theta, \pi/2)$ and $H_m(\theta, \pi/2)$.

The amplitude function $T(\lambda/2)$ has the form:

$$T\left(\frac{\lambda}{2}\right) = - \left[\frac{S_a\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \pm S_b\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \pm \frac{1}{2} \left[S_b\left(\frac{\lambda}{2}, 0\right) - S_b\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right] + j \operatorname{Im} \frac{1}{2} \left[S_a\left(\frac{\lambda}{2}, 0\right) - S_a\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right]}{\Psi_{av} + C_a\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \pm C_b\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) + E_a\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \pm E_b\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right)} \right], \quad (93a)$$

$$T^{(0)}\left(\frac{\lambda}{2}\right) = -0.216 + j0.274, \quad T^{(1)}\left(\frac{\lambda}{2}\right) = -0.177 + j0.066. \quad (93b)$$

THE TWO-ELEMENT ARRAY

A valuable insight into the properties of arrays in their dependence upon the distributions of current in the several elements, may be obtained from a study of an array of two full-wave dipoles. As the simplest special case of the circular array, it may be analyzed accurately in terms of the zero phase-sequence or symmetrical case in which the two elements are driven by equal voltages in phase, and the first phase sequence or antisymmetrical case in which the two driving voltages are equal but 180 degrees out of phase. Taken separately, these phase sequences correspond to the two-element broadside and the two-element bidirectional end-fire arrays, in each of which the distributions of current along the elements are necessarily the same. By superposition, the general case of arbitrary driving voltages may be obtained including, among others, the case having one element parasitic, with an arbitrary tuning or loading impedance at its center, and the case having two-element unidirectional end-fire array or couplet, where b is the distance between the elements and $\beta = 2\pi/\lambda$.

For two identical elements for each of which $\beta h = \pi$, $\Omega = 10$ and $\beta b \geq 1$, Ψ_{dR} and Ψ_{dI} are the same as for an isolated element; they are given by (80b) and (80c). Ψ_{dV} and Ψ_{dZ} may be determined from the following formulas; the upper sign is for $\Psi^{(0)}$, and the lower is for $\Psi^{(1)}$. The numerical values are for $\beta b = 1.5$.

$$\Psi_{dV} = \frac{1}{2} \left\{ \left[C_a\left(\frac{\lambda}{2}, 0\right) - C_a\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right] \pm \left[C_b\left(\frac{\lambda}{2}, 0\right) - C_b\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right] + \left[E_a\left(\frac{\lambda}{2}, 0\right) - E_a\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right] \pm \left[E_b\left(\frac{\lambda}{2}, 0\right) - E_b\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right] \right\} \quad (91a)$$

$$\Psi_{dV}^{(0)} = 7.848 - j3.939, \quad \Psi_{dV}^{(1)} = 7.352 - j0.661 \quad (91b)$$

$$\Psi_{dZ} = \pm \frac{1}{2} \left[S_b\left(\frac{\lambda}{2}, 0\right) - S_b\left(\frac{\lambda}{2}, \frac{\lambda}{2}\right) \right] \quad (92a)$$

$$\Psi_{dZ}^{(0)} = -0.245 - j0.524; \quad \Psi_{dZ}^{(1)} = 0.245 + j0.524. \quad (92b)$$

With these values the quasi-zero-order currents (in amperes when V_0 is in volts) on the two antennas in the two phase sequences are given by the following formulas when $\beta b = 1.5$:

$$I_{2z}^{(0)} = I_{1z}^{(0)} = V^{(0)} \{ 0.783(\cos \beta z + 1) - j[2.805 \sin \beta z - 0.617(\cos \beta z + 1)] \} \times 10^{-3}. \quad (94a)$$

$$-I_{2z}^{(1)} = I_{1z}^{(2)} = V^{(1)} \{ 0.189(\cos \beta z + 1) - j[2.805 \sin \beta z - 0.506(\cos \beta z + 1)] \} \times 10^{-3}. \quad (94b)$$

These currents are shown in Fig. 10 in the form $I_z = I_z'' + jI_z'$ where I_z'' is in phase and I_z' in phase quadrature with V_0 . The corresponding driving-point admittances and impedances are

$$Y^{(0)} = (1.566 + j1.234) \times 10^{-3} \text{ mhos}, \quad Y^{(1)} = (0.378 + j1.012) \times 10^{-3} \text{ mhos}, \quad (95)$$

$$Z^{(0)} = 394 - j310 \text{ ohms}, \quad Z^{(1)} = 324 - j867 \text{ ohms}. \quad (96)$$

The self- and mutual impedances are defined by

$$Z_{s1} = \frac{1}{2}(Z^{(0)} + Z^{(1)}) = 359 - j588 \text{ ohms}, \quad (97)$$

$$Z_{12} = \frac{1}{2}(Z^{(0)} - Z^{(1)}) = 70 - j278 \text{ ohms}. \quad (98)$$

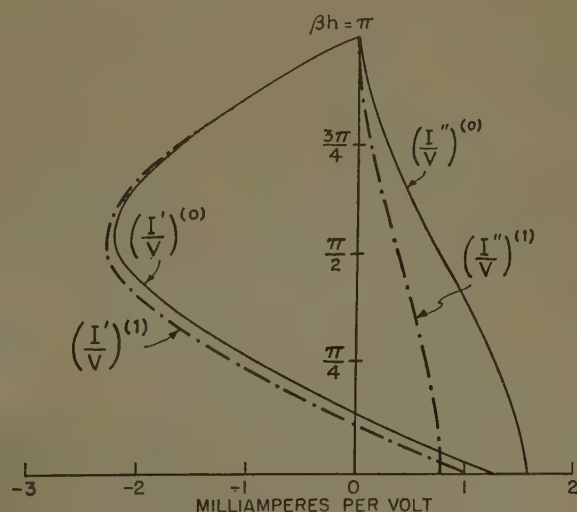


Fig. 10—Zero- and first-phase-sequence currents on two-element array. $\Omega = 10$, $\beta b = 1.5$.

If antenna 1 is driven and antenna 2 is an unloaded parasitic, the currents are obtained by setting $V_{10} = V^{(0)} + V^{(1)}$ with $V^{(1)} = V^{(0)}$. It follows that $V_{20} = V^{(0)} - V^{(1)} = 0$. The currents are

$$I_{1z} = I_z^{(0)} + I_z^{(1)} = V_{10} \{ 0.486(\cos \beta z + 1) - j[2.805 \sin \beta z - 0.566(\cos \beta z + 1)] \} \times 10^{-3}, \quad (99a)$$

$$I_{2z} = I_z^{(0)} - I_z^{(1)} = V_{10}(0.287 + j0.055(\cos \beta z + 1)). \quad (99b)$$

The corresponding driving-point admittance and impedance are

$$Y_{10} = (0.972 + j1.33) \times 10^{-3} \text{ mhos},$$

$$Z_{10} = 436 - j508 \text{ ohms}. \quad (100)$$

The currents in the driven antenna and the untuned parasitic element when $\beta b = 1.5$ are shown in Fig. 11(a). They are seen to differ greatly, in both distribution and magnitude. Indeed, contributions to the far-zone field by the currents in the parasitic element will be insignificant, and the horizontal field pattern must be almost a circle. Note that this behavior is entirely different from that which it would be if the two elements were half-wave instead of full-wave dipoles. In the former case, the current in the parasitic element would be comparable in magnitude and essentially the same in distribution as that in the driven element. The reason for this difference is to be found in the fact that the half-wave elements are nearly tuned to resonance, whereas the full-wave elements are detuned or near antiresonance. This condition can be changed by inserting a lumped reactance (or an equivalent section of transmission line) in series with the full-wave parasitic antenna at its center and by tuning this reactance to make the entire circuit resonant. The analytical steps and the theoretical distributions of current are readily obtained.

The currents in a driven full-wave dipole and an identical parasitic element with a lumped reactance at its center may be expressed in the form

$$I_{1z} = V^{(0)}f_z^{(0)} + V^{(1)}f_z^{(1)} = V_{10}S_z + V_{20}D_z \quad (101a)$$

$$I_{2z} = V^{(0)}f_z^{(0)} - V^{(1)}f_z^{(1)} = V_{10}D_z + V_{20}S_z \quad (101b)$$

where $F_z^{(0)}$ is the coefficient of $V^{(0)}$ in (94a) and $f_z^{(1)}$ is the coefficient of $V^{(1)}$ in (94b); $S_z = \frac{1}{2}[f_z^{(0)} + f_z^{(1)}]$ and $D_z = \frac{1}{2}[f_z^{(0)} - f_z^{(1)}]$. If z is set equal to zero in (101a) and (101b), and the lumped impedance Z_2 is introduced by setting $V_{20} = -I_{20}Z_2$, it is readily shown that

$$V_{20} = -V_{10}Z_2D_0/(1 + Z_2S_0). \quad (102)$$

With this value of V_{20} , (101a) and (101b) become

$$I_{1z} = V_{10}[S_z - D_0D_zZ_2/(1 + Z_2S_0)] \quad (103a)$$

$$I_{2z} = V_{10}[D_z - D_0S_zZ_2/(1 + Z_2S_0)]. \quad (103b)$$

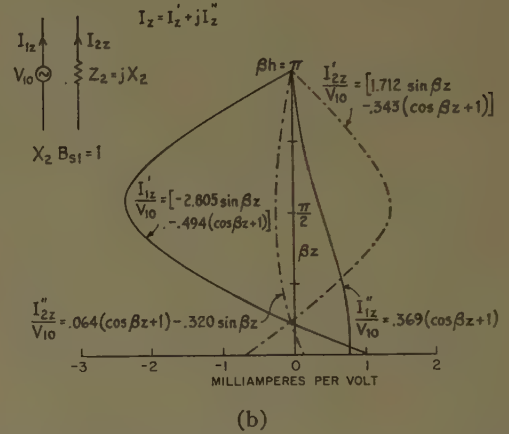
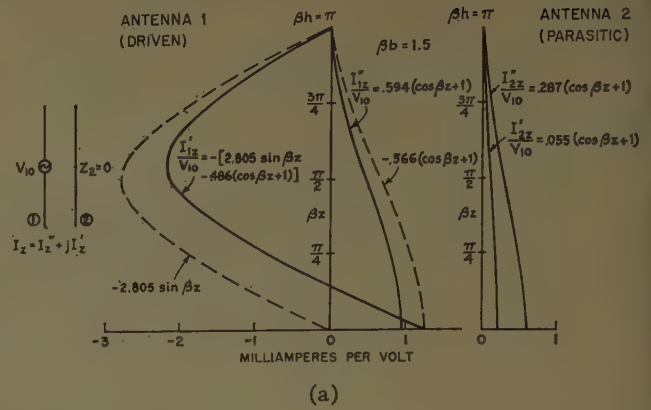


Fig. 11—(a) Currents on full-wave antenna with untuned parasitic; $\Omega = 10$; (b) currents on full-wave antenna with tuned parasitic; $\Omega = 10$.

The parasitic element may now be tuned to resonance by setting $Z_2 = jX_2$ and by choosing X_2 so that X_2B_{s1} is the imaginary part of $S_0 = Y_{s1} = G_{s1} + jB_{s1}$. With this choice, (103a) and (103b) reduce to:

$$I_{1z} = V_{10}[S_z - D_0D_z/G_{s1}] \quad (104a)$$

$$I_{2z} = V_{10}[D_z - D_0S_z/G_{s1}] \quad (104b)$$

where G_{s1} is the real part of S_0 . Numerical values of the currents when $\beta b = 1.5$ are

$$I_{1z} = V_{10} \{ 0.369(\cos \beta z + 1) - j[2.805 \sin \beta z - 0.494(\cos \beta z + 1)] \} \times 10^{-3} \quad (105a)$$

$$I_{2z} = V_{10} \{ [-0.064(\cos \beta z + 1) - 0.320 \sin \beta z + j[1.712 \sin \beta z - 0.343(\cos \beta z + 1)]] \} \times 10^{-3}. \quad (105b)$$

These currents are shown in Fig. 11(b). They are seen to be more nearly alike in both distribution and amplitude than those in Fig. 11(a) so that the horizontal field pattern of the full-wave array with a tuned parasitic element should approximate at least roughly that of the half-wave array with an unloaded parasitic element.

The most interesting two-element array is the coupler in which the distance between elements is $\lambda/4$ and the currents at the driving points are equal in amplitude but differ in phase by a quarter period. Thus, let $\beta b = 1.5$ (strictly, this should be $\pi/2$) and let $I_{20} = jI_{10}$. The cur-

rents are readily shown to be given by (101a) and (101b), with

$$V_{20} = jV_{10} \left[\frac{S_0 + jD_0}{S_0 - jD_0} \right] = -0.966 + j1.267$$

$$= 1.59e^{-j146.3^\circ} \quad (106)$$

With this value, the explicit formulas for the current in an array with $\beta h = \pi$, $\Omega = 10$ are

$$I_{1z} = V_{10} \{ [0.129(\cos \beta z + 1) - j[2.805 \sin \beta z - 0.884(\cos \beta z + 1)]] \} \times 10^{-3} \quad (107a)$$

$$I_{2z} = V_{20} \{ [0.400(\cos \beta z + 1) - j[2.805 \sin \beta z - 0.392(\cos \beta z + 1)]] \} \times 10^{-3} \quad (107b)$$

In order to obtain expressions for the current that are comparable from the point of view of maintaining an electromagnetic field, it is necessary to use the same reference for amplitude and phase. If I_{2z} is referred to V_{10} instead of V_{20} , the following formula is obtained in place of (107b):

$$I_{2z} = V_{10} \{ [3.554 \sin \beta z - 0.884(\cos \beta z + 1) + j[2.710 \sin \beta z + 0.129(\cos \beta z + 1)]] \} \times 10^{-3} \quad (107c)$$

The corresponding driving point admittances and impedances are

$$Y_{10} = (0.258 + j1.768) \times 10^{-3} \text{ mhos},$$

$$Y_{20} = (0.801 + j0.784) \times 10^{-3} \text{ mhos} \quad (108a)$$

$$Z_{10} = 80.8 - j554 \text{ ohms}, \quad Z_{20} = 638 - j624 \text{ ohms}. \quad (108b)$$

The ratio of the power supplied to antenna 1 to that supplied to antenna 2 is $|V_{20}|^2 G_{20} / |V_{10}|^2 G_{10} = 7.9$. The currents represented by (107a) and (107b) are shown in the upper diagram in Fig. 12 in the form I_{1z}/V_{10} and I_{2z}/V_{20} . The distribution of I_{2z}/V_{10} is shown in the bottom diagram in Fig. 12. It is seen to differ greatly from I_{1z}/V_{10} (shown in the upper graph) even though the input currents at $z=0$ satisfy the assigned relation, $I_{20} = jI_{10}$.

The radiation field of the full-wave couplet may be expressed as follows:

$$E_{\theta}^r = E_{\theta 1}^r + E_{\theta 2}^r$$

$$= K[A_1 e^{j(\beta b/2) \cos \Phi} + A_2 e^{-j(\beta b/2) \cos \Phi}], \quad (109)$$

where

$$A_1 = V_{10}[(0.129 + j0.884)G_m(\theta, \pi) - j2.805F_m(\theta, \pi)], \quad (110a)$$

$$A_2 = V_{10}[(-0.884 + j0.129)G_m(\theta, \pi) + (3.554 + j2.710)F_m(\theta, \pi)], \quad (110b)$$

and where K is a constant. Note that in the equatorial plane, $\theta = \pi/2$, and $G_m(\pi/2, \pi) = \pi$, $F_m(\pi/2, \pi) = 2$. The field pattern calculated from the magnitude of (109) with (110a) and (110b) for the couplet of full-wave elements is shown in Fig. 13 together with the corresponding pattern for the ideal couplet with identical distributions

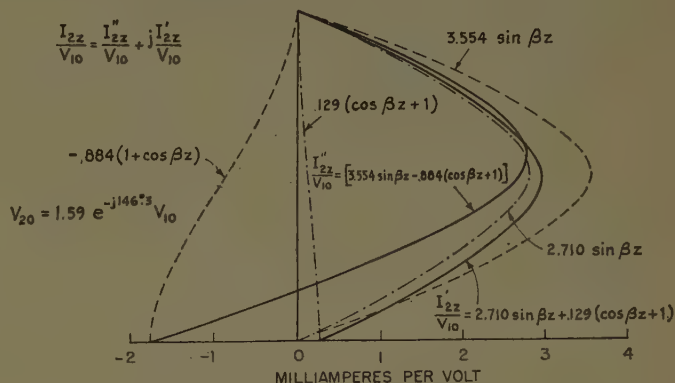
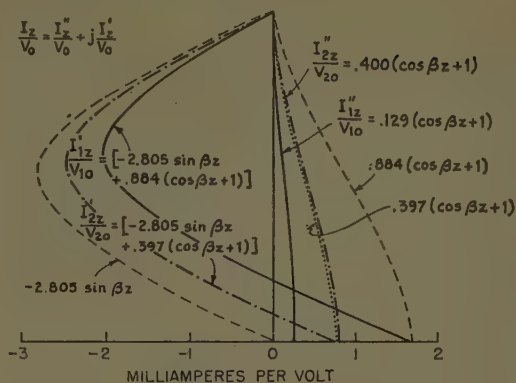


Fig. 12—Currents in full-wave couplet:
 $I_{20} = jI_{10}$; $\beta b = 2\pi b/\lambda = 1.5$; $\Omega = 10$.

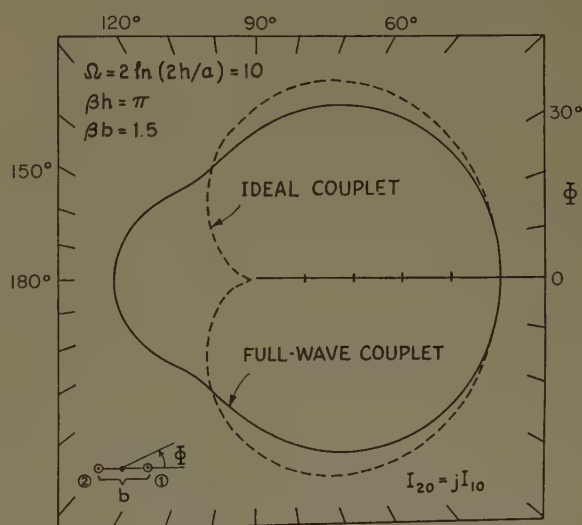


Fig. 13—Horizontal pattern of full-wave couplet with $I_{20} = jI_{10}$.

of current in the two elements. (This latter is quite closely approximated by the pattern of a couplet of half-wave elements.) Both patterns are normalized to unity at $\Phi = 0$. It is seen that the deep minimum at $\Phi = 180^\circ$ in the ideal pattern (this would be a null if $\beta b = \pi/2$ had been used instead of $\beta b = 1.5$) is replaced by a minor maximum with an amplitude that is about one half that of the principal maximum at $\Phi = 0$. Thus, the characteristic property of the ideal couplet of providing a null in one direction does not exist in actual couplets when $\beta h = \pi$ or, in fact, for any other value of βh that is

not near $\pi/2$ or that is not an odd multiple thereof. Significantly, this makes the cardioid pattern of the half-wave couplet a relatively narrow-band property!

The question now arises as to whether it is possible to make the two distributions of current alike by tuning each element to self-resonance, as was done in the case of the parasitic element, and so, perhaps, to regain the ideal field pattern. This is readily investigated by proceeding from (101a) and (101b) and by setting $V_{10} = V_1^e - I_{10}Z_1$, $V_{20} = V_2^e - I_{20}Z_2$. V_1^e and V_2^e are the EMF's of the generators with series impedances Z_1 and Z_2 which may be pure tuning reactances. With this substitution and some algebraic manipulation, the following expressions for the current are obtained:

$$I_{1z} = \frac{1}{\Delta} \{ V_1^e [(1 + S_0 Z_2) S_z - D_0 Z_2 D_z] + V_2^e [-D_0 Z_1 S_z + (1 + S_0 Z_1) D_z] \} \quad (111a)$$

$$I_{2z} = \frac{1}{\Delta} \{ V_1^e [(1 + S_0 Z_2) D_z - D_0 Z_2 S_z] + V_2^e [-D_0 Z_1 D_z + (1 + S_0 Z_1) S_z] \} \quad (111b)$$

where

$$\Delta = (1 + S_0 Z_1)(1 + S_0 Z_2) - D_0^2 Z_1 Z_2. \quad (111c)$$

If the condition $I_{20} = jI_{10}$ is now imposed and the antennas are individually tuned to self-resonance by requiring that $Z_1 = Z_2 = jX = j/\text{Im } S_0$, the EMF's must satisfy the relation $V_2^e = jCV_1^e$, with

$$C = \frac{[S_0 + jD_0 + Z_2(S_0^2 - D_0^2)]}{[S_0 - jD_0 + Z_1(S_0^2 - D_0^2)]} = \frac{[(S_0 + jD_0) \text{Re } S_0 - D_0(D_0 + jS_0)]}{[(S_0 - jD_0) \text{Re } S_0 - D_0(D_0 - jS_0)]}. \quad (112)$$

The currents in the same full-wave couplet previously considered, but with the elements tuned to self-resonance, are given by

$$I_{1z} = V_1^e \{ [(1 - 0.676 \sin \beta z + 0.338(\cos \beta z + 1))] + j[4.232 \sin \beta z - 1.410(\cos \beta z + 1)] \} \times 10^{-3} \quad (113a)$$

$$I_{2z} = V_2^e \{ [(-3.334 \sin \beta z + 0.669(\cos \beta z + 1))] - j[1.399 \sin \beta z + 0.280(\cos \beta z + 1)] \} \times 10^{-3} \quad (113b)$$

where $V_2^e = jCV_1^e$ with $C = 1.615 + j1.182 = 2.00 e^{j36.2}$. It follows that

$$I_{2z} = V_1^e \{ [-3.733 \sin \beta z + 1.411(\cos \beta z + 1)] - j[6.203 \sin \beta z - 0.339(\cos \beta z + 1)] \} \times 10^{-3}. \quad (113c)$$

The driving-point admittances in mhos and impedances in ohms are

$$Y_{10} = (0.676 - j2.420) \times 10^{-3},$$

$$Y_{20} = (1.338 - j0.560) \times 10^{-3} \quad (114)$$

$$Z_{10} = 80.8 + j336, \quad Z_{20} = 638 + j266. \quad (115)$$

Note that the resistances in (115) are the same as those in (108b) but that the reactances differ by the magnitude of the tuning impedance $Z_1 = Z_2 = jX = j/\text{Im } S_0 = j890$ ohms. The distributions I_{1z}/V_{10} and I_{2z}/V_{20} are shown at the top in Fig. 14. They depart considerably from the distributions in Fig. 12 for the same antennas when untuned. Significantly, the ratio of the power in antenna 2 to that in antenna 1 differs negligibly from the value 7.9 obtained for the untuned array

The far-zone field of the couplet with the currents given by (113a) and (113c) is readily calculated from (109), with

$$A_1 = V_1^e [(0.338 - j1.410)G_m(\theta, \pi) - (1.676 - j4.232)F_m(\theta, \pi)] \times 10^{-3} \quad (116a)$$

$$A_2 = V_1^e [(1.410 + j0.338)G_m(\theta, \pi) - (3.733 + j6.203)F_m(\theta, \pi)] \times 10^{-3}. \quad (116b)$$

However, for a convenient comparison of field patterns, it is first desirable to adjust the driving voltage V_1^e in (116a) and (116b) so that the same total power is supplied to the tuned array as to the untuned one with the currents given in (107a) and (107c). The power transferred to, and radiated from, the untuned array is

$$P_U = \frac{1}{2} |V_{10}|^2 [G_{10} + |V_{20}/V_{10}|^2 G_{20}] = \frac{1}{2} |V_{10}|^2 [0.258 + (1.59)^2 \times 0.801] \times 10^{-3} = |V_{10}|^2 \times 1.13 \times 10^{-3}; \quad (117)$$

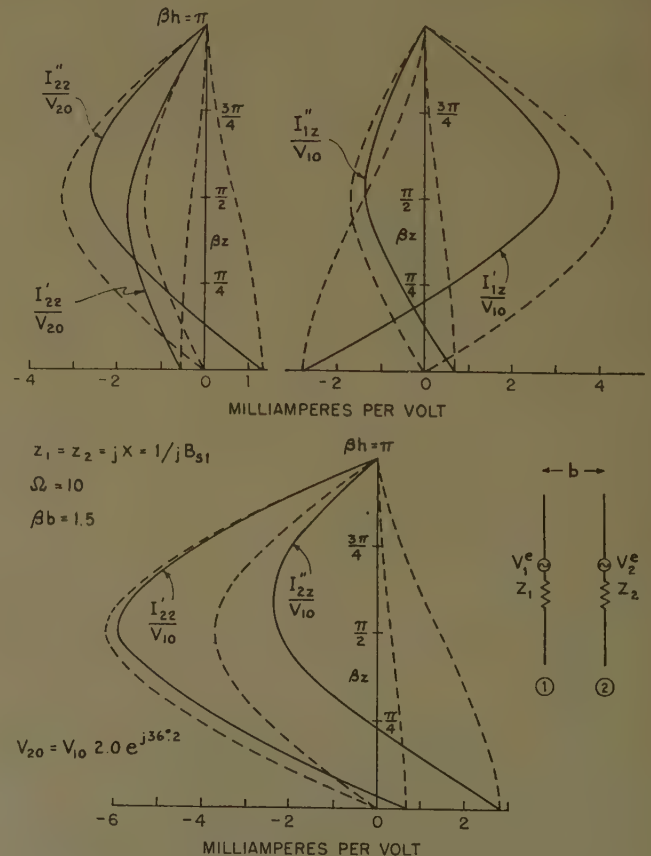


Fig. 14—Currents on elements of tuned full-wave couplet; $I_{20} = jI_{10}$.

the power transferred to, and radiated from, the tuned array is

$$P_T = \frac{1}{2} |V_1^e|^2 [0.676 + (2.0)^2 \times 1.338] \times 10^{-3} \\ = |V_1^e|^2 \times 3.014 \times 10^{-3} \quad (118)$$

These powers are in watts when V_{10} and V_1^e are in volts. For equal powers to the two couplets, the ratio of their driving voltages must be $|V_1^e|/|V_{10}| = 1.635$. When the magnitudes of the driving voltages are adjusted to this ratio, the horizontal field pattern is found to *differ negligibly* from that shown in Fig. 13 for the same couplet when the elements are driven by impedanceless generators. Thus, the tuning of the elements serves to alter the input reactances and the distributions of current but not the magnitude of the electric field. Evidently, tuning the elements of a full-wave couplet has quite a different effect on the field pattern from tuning the parasitic element in an array of two full-wave elements of which only one is driven. This is not surprising if it is recalled that in the former the relative amplitudes and phases at the centers of the two antennas are assigned, whereas in the latter they are not. The important conclusion is that the horizontal field pattern of the full-wave couplet differs greatly from that of the ideal or half-wave couplet whether the individual elements are tuned or not. The useful sharp minimum or null at $\Phi = 0$, characteristic of the cardioid, is not obtained.

ARRAYS OF MANY ELEMENTS

The determination of the distributions of current, the input admittances, and the radiation fields of arbitrarily driven N -element circular arrays may be carried out in a manner paralleling that illustrated for the simple two-element array. The formulation for each of the N -phase sequences has been given, and any arbitrary set of driving conditions may be obtained by a suitable superposition of these. The analysis of specific arrays of this type is reserved for a subsequent paper. However, it is clear from the discussion of the two-element array that when elements appreciably longer than a half-wavelength are involved, the conventional field pattern obtained by the product of array factors and field factors cannot yield the correct result. The entire minor lobe structure is affected by differences in the distributions of current in the several elements. In particular, nulls will disappear and possibly even be changed to relative maxima. When an array is used over a band of frequencies, the details of the minor lobe structure may be expected to vary with frequency from a pattern closely resembling that given by the conventional product of field factor and array factors when the dipoles are near a half-wavelength long, to a pattern dif-

fering greatly from this ideal at frequencies for which the length of the dipoles approaches the full wavelength even if the spacing of the elements is kept constant in terms of the wavelength.

In this paper only the circular array has been analyzed. However, the procedure for determining the currents in simple trigonometric form may be extended, with certain modifications and approximations, to parallel arrays in general. A discussion of this problem is also reserved for a subsequent paper.

In conclusion, it may be stated that for the arbitrarily-driven circular array a solution has been obtained for the distributions of current in the several elements in a form that is both reasonably accurate and sufficiently simple to permit the evaluation of the far-zone field in closed form. The field of each element is represented by two field factors with complex coefficients that depend on the location of the element in the array and on the individual driving conditions. The conventional product form of the far-zone field of an array is inadequate except for arrays (*e.g.*, half-wave dipoles) in which the *distributions* of driven and induced currents are essentially alike.

The statement that scattering and diffraction by the elements of an array is responsible for the failure of the product formula for the far-zone field is misleading. Actually, the entire field of an array may be ascribed to scattering and diffraction including, especially, the major and minor lobe structure and the nulls that occur in the ideal patterns. Scattering by an array may also be described as reradiation by the induced currents. In special cases, the distributions of these induced currents may be so nearly alike that a single field factor adequately characterizes the contribution of each; more generally, the distributions are different, in which case the assumption that they are alike naturally leads to incorrect results. Obviously, it is not only the *differences* between the field pattern in the ideal special case, and in more general ones, that must be ascribed to scattering by the array, but the entire phenomenon of re-enforcement and cancellation of the fields maintained by the currents in the elements. When these currents are identical, the product formula correctly describes the scattering and diffraction by the array; when they are not, the scattering and diffraction by the array are more complicated, but not fundamentally different, and the product formula is strictly not applicable.

Differences in the distributions of current in the elements of an array are primarily responsible for the failure of the product formula. Such differences always exist when the distributions of induced and forced currents are not alike and the electrical environments of the individual elements are not the same.

A New Method of Near Field Analysis*

R. C. HANSEN† AND L. L. BAILIN‡

Summary—The problem of the accurate determination of the total electromagnetic field produced by a current distribution on a circular disc type aperture is examined. For comparison of the near and Fresnel region solutions, an exact Maxwellian integral is used as the starting point. Aperture and space coordinates are separated by the use of Bessel function addition theorems, allowing the field to be expressed as a sum of terms containing radial and angular integrals. The radial integrals involve only the disc (aperture) distribution and coordinates while the angular integrals involve only the elevation angle. The distance from origin appears implicitly in the series coefficients. These integrals are suitable for high speed computations and have been evaluated for a number of cases. The results are compared with the small angle and general Fresnel formula values for various observation angles, distances from the aperture, and aperture distributions.

I. INTRODUCTION

CALCULATION of the total field radiated by a current distribution on an antenna surface is an old problem with an extensive literature.¹ Information on the field between the antenna and the far field region is usually given in terms of approximate small angle Fresnel formulas. However, recent advances in antenna and systems techniques have made an accurate knowledge of this field imperative. Some of the important problems requiring such a knowledge are presented.

- 1) What are the limits of far field synthesis of very low sidelobe antennas?
- 2) How do "optimum" low sidelobe designs deteriorate in the Fresnel region as compared to other low sidelobe designs?
- 3) What is the range of validity in angle and distance of the small angle Fresnel approximation and of the general Fresnel formula?
- 4) What is the effect of the edge line integral?
- 5) How does the maximum power density vary with distance?
- 6) In the calculation of far field characteristics from limited Fresnel region measurements, what measured quantities are most sensitive?

In order to provide the answers to some of these questions, this paper critically examines the degree of validity of the various classical formulations of radiation fields by comparing them with the exact field. The exact

field is first formulated in terms of definite integrals which are amenable to evaluation by a high speed computer. The accurate data obtained are then used to evaluate the effects of the near field and the edge line integral contributions which are completely ignored in the classical formulations.

Because the flat circular surface, commonly called a circular aperture, possesses unique properties not shared by rectangular surfaces, it is taken as the basis for this study. This surface furnishes an approximation to a parabolic dish antenna, and also to a planar array of radiating elements. The exact field due to a current distribution (including edge contributions) on one side of a flat circular perfectly conducting dish is cast into a form suitable for numerical evaluation. Results of the computations are presented in Section III. In Section IV both small angle and general Fresnel formulas are derived, and the field plots given. Axial power variations of the several results are compared in Section V. Because of differences in common usage, the near, Fresnel, Fraunhofer, and far field regions, are defined in the Appendix.

II. EXACT FIELD FORMULATION

For a field description which includes near field terms ($1/r^2$ and higher in aperture coordinates) and edge line integrals, the integrated wave equation formulation is a suitable starting point. The usual starting point, a scalar Kirchhoff type integral, omits such contributions.

In the geometry of Fig. 1, a linear current density polarized in the x direction flows on the circular surface in the xy plane. Circular and other polarizations can be calculated by appropriate combination of linearly polarized solutions. Cylindrical coordinates ($\rho, \beta, z=0$) are used for the current J and the integration point Q on the disc. Only rotationally symmetric current variations in which the current amplitude is a function $f(k\rho)$ and independent of β will be considered.

For a point of observation P whose position in space is given by the conventional spherical coordinates (R, θ, ϕ), the integration of the wave equation given by Silver² yields

$$E = \frac{-j\eta}{4\pi} \int_A [(J \cdot \nabla_{kr}) \nabla_{kr} + k^2 J] \psi dA, \quad (1)$$

$$H = \frac{1}{4\pi} \int_A (J \times \nabla_{kr}) \psi dA, \quad (2)$$

* This work has been partially supported by the AF Cambridge Research Center, Air Res. and Dev. Command, under Contract AF 19(604)-3508.

† Microwave Laboratory, Hughes Aircraft Co., Culver City, Calif.

‡ Dept. of Electrical Engrg., University of Southern California, Los Angeles, Calif.

¹ C. J. Bouwkamp, "Diffraction theory," *Repts. Progr. Phys.*, vol. 17, pp. 35-100; 1954.

² S. Silver, "Microwave Antenna Theory and Design," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 12, p. 160; 1949.

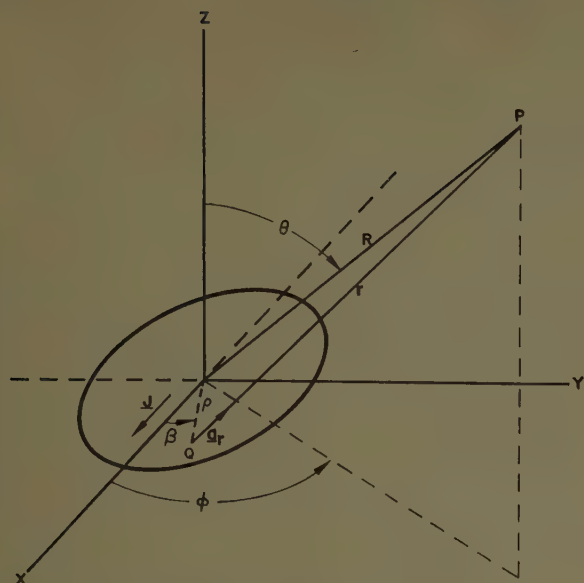


Fig. 1—Aperture and field coordinates.

where $e^{-j\omega t}$ is suppressed, $\eta = 120\pi$, $k = 2\pi/\lambda$, $\psi = e^{jkr}/kr$, and

$$r^2 = R^2 + \rho^2 - 2R\rho \sin \theta \cos (\phi - \beta). \quad (3)$$

Note that the integral in (1) contains terms derived from a line integral of charges at the edge of the aperture. For aperture distributions $J(k\rho)$ with discontinuities at the edge, these terms are necessary to produce a Maxwellian field. The current on the back side of the dish is assumed to be zero. An excellent discussion of the Kottler edge integral formulation is given by Bouwkamp.³

The restricted vector operators in (1) and (2) are expanded

$$\nabla_{kr}\psi = \left(j + \frac{1}{kr}\right)\mathbf{a}_r, \quad J \times \nabla_{kr}\psi = \left(j + \frac{1}{kr}\right)J \times \mathbf{a}_r\psi, \quad (4)$$

$$(J \cdot \nabla_{kr})\nabla_{kr}\psi$$

$$= \left[\left(-1 + \frac{3j}{kr} + \frac{3}{k^2r^2}\right)(J \cdot \mathbf{a}_r)\mathbf{a}_r - \frac{1}{kr}\left(j + \frac{1}{kr}\right)J \right]\psi \quad (5)$$

with \mathbf{a}_r a unit vector from Q to P . The vector \mathbf{a}_r can be expressed in terms of spherical components as

$$\begin{aligned} \mathbf{a}_r = \mathbf{a}_R \left[\frac{R}{r} - \frac{\rho}{r} \sin \theta \cos (\phi - \beta) \right] \\ - \mathbf{a}_\theta \frac{\rho}{r} \cos \theta \cos (\phi - \beta) \\ + \mathbf{a}_\phi \frac{\rho}{r} \sin (\phi - \beta). \end{aligned} \quad (6)$$

For a linear current,

$$J = f(k\rho)[\mathbf{a}_R \sin \theta \cos \phi + \mathbf{a}_\theta \cos \theta \cos \phi - \mathbf{a}_\phi \sin \phi] \quad (7)$$

³ Bouwkamp, *op. cit.*, Sec. V.

and the scalar product is

$$(J \cdot \mathbf{a}_r) = f(k\rho) \left[\frac{R}{r} \sin \theta \cos \phi - \frac{\rho}{r} \cos \beta \right]. \quad (8)$$

Now the total field expression can be written, but only the E_θ component will be discussed here. (Details of other components will be described in a forthcoming Hughes Aircraft Company Scientific Report.) Thus, from (1)

$$\begin{aligned} E_\theta = \frac{-j\eta}{4\pi} \int_0^{2\pi} \int_0^{ka} \left[\left(1 - \frac{3j}{kr} - \frac{3}{k^2r^2}\right) \right. \\ \cdot \left(\frac{R}{r} \sin \theta \cos \phi - \frac{\rho}{r} \cos \beta \right) \frac{\rho}{r} \cos \theta \cos (\phi - \beta) \\ \left. + \left(1 - \frac{j}{kr} - \frac{1}{k^2r^2}\right) \cos \theta \cos \phi \right] \psi k\rho f(k\rho) d(k\rho) d\beta, \quad (9) \end{aligned}$$

where a is the radius of the disc.

At this point several schemes for the numerical evaluation of these integrals can be considered. Integration in closed form is impossible with presently defined functions, since, as will appear later, only the simplest term in (9) can be written in terms of Lommel functions. The remaining terms are not even amenable to solution by series expansion techniques since they involve the complicated distance r which is a function of both P and Q . A direct numerical integration suffers from the major difficulty that a new integration is necessary for each point P , again due to the complexities of r . If various aperture distributions are to be considered, it is desirable that a partial separation of r into the aperture and space coordinates be effected. To this end functions of r are expressed as separable functions of R and ρ . This approach is the essence of the method used here.

As a first step in the separation the integrand of (9) is written in terms of spherical Bessel functions as⁴

$$\begin{aligned} E_\theta = \frac{\eta}{4\pi} \iint \left\{ h_2^{(2)}(kr) \left(\frac{R}{r} \sin \theta \cos \phi - \frac{\rho}{r} \cos \beta \right) \frac{\rho}{r} \right. \\ \cdot \cos \theta \cos (\phi - \beta) + \left[\frac{h_1^{(2)}(kr)}{kr} - h_0^{(2)}(kr) \right] \\ \cdot \cos \theta \cos \phi \left. \right\} f(k\rho) k\rho dk\rho d\beta. \end{aligned} \quad (10)$$

Then, the well-known Bessel function addition theorems⁵ are employed and an infinite series is obtained from each Bessel function in (10). To save space the steps in combining and renumbering the three series are omitted and the result is given as

⁴ E. P. Adams, "Smithsonian Mathematical Formulae and Tables of Elliptic Functions," Smithsonian Institution, Washington, D. C., p. 203; 1947.

⁵ P. M. Morse and H. Feshbach, "Methods of Theoretical Physics," McGraw-Hill Book Co., Inc., New York, N. Y., vol. 2, p. 1574; 1953.

$$E_\theta = \frac{\eta}{4\pi} \sum_{N=0}^{\infty} (2N+1) h_N^{(2)}(kR) \int_0^{2\pi} \int_0^{ka} j_N(k\rho) \left\{ \left(\frac{\sin \theta \cos \phi}{k\rho} - \frac{\cos \beta}{kR} \right) \frac{\cos \theta \cos(\phi - \beta)}{kR} T_{N-2}(\kappa) \right. \\ \left. + \frac{\cos \theta \cos \phi}{k\rho kR} T_{N-1}(\kappa) - \cos \theta \cos \phi T_N(\kappa) \right\} f(k\rho) k\rho d(k\rho) d\beta, \quad (11)$$

where the argument $\kappa = \sin \theta \cos(\phi - \beta)$, j_n and $h^{(2)}$ are spherical Bessel functions, and the T_p^m are Gegenbauer polynomials for $p > 0$ and zero for $p \leq 0$. Eq. (11) provides the desired separation of the functions of R and ρ . Thus if we rewrite (11) using a shorthand notation which defines the radial integrals

$$B_1(N) = \int_0^{ka} f(k\rho) j_N(k\rho) k\rho d(k\rho), \\ B_2(N) = \int_0^{ka} f(k\rho) j_N(k\rho) d(k\rho), \quad (12)$$

and the angular integrals

$$A_1(N) = \int_0^{\pi/2} T_N(\sin \theta \cos \beta) d\beta \quad \text{for } N \text{ even}, \\ A_2(N) = \int_0^{\pi/2} T_N^1(\sin \theta \cos \beta) d\beta \quad \text{for } N \text{ even}, \\ A_4(N) = \int_0^{\pi/2} T_N^2(\sin \theta \cos \beta) \cos \beta d\beta \quad \text{for } N \text{ odd}, \\ A_5(N) = \int_0^{\pi/2} T_N^2(\sin \theta \cos \beta) \cos^2 \beta d\beta \quad \text{for } N \text{ even}, \quad (13)$$

then the field can be simply written as

$$E_\theta = \frac{\eta \cos \theta \cos \phi}{\pi} \sum_{\substack{N=0 \\ \text{even}}}^{\infty} (2N+1) h_N^{(2)}(kR) \\ \cdot \left[-A_1(N) B_1(N) + \frac{1}{kR} A_2(N-1) B_2(N) \right. \\ \left. - \frac{1}{k^2 R^2} A_5(N-2) B_1(N) \right] \\ + \frac{\eta \sin \theta \cos \theta \cos \phi}{\pi kR} \\ \cdot \sum_{\substack{N=1 \\ \text{odd}}}^{\infty} (2N+1) h_N^{(2)}(kR) A_4(N-2) B_2(N). \quad (14)$$

In obtaining (14) from (11), it can be shown that the properties of the Gegenbauer polynomials reduce the summations to the values indicated, the range of β from 2π to $\pi/2$, and the argument κ to $\sin \theta \cos \beta$ as given in (13) and (14).

Eq. (14) is in a form which is convenient for numerical evaluation, since the radial integrals $B(N)$ contain the aperture function $f(k\rho)$, and they do not depend

upon the observation point P . Also, the angular integrals $A(N)$ do not involve the aperture size or distribution; in fact the only parameter is the elevation angle θ . The cost of this type of separation is of course an infinite series and a number of integrals that must be computed. The flexibility obtained, and the avoidance of "starting over" each time a parameter changes, is well worth the cost. This technique appears to make near field calculations feasible for the first time. Methods for computation of the integrals have been discussed in an earlier article.⁶ The convergence of (14) is produced by the $B(n)$ integrals and is independent of the distance of the observation point from the origin. The next section will present the computed results.

III. COMPUTED FIELD DISTRIBUTIONS

Computations of the angular and radial integrals were made on an IBM 704, for ranges of $\theta = 0$ (2.5) 90° , and $ka = 10\pi, 20\pi, 40\pi$; only the 10π results are presented here. Four distribution functions were used: uniform, 25-, 30-, 35-db Taylor optimum low sidelobe designs.^{7,8} Table I gives the pertinent data for these four distributions. Although dish aperture distributions in the literature are usually of the $(1 - k^2 \rho^2)^N$ type, these types are neither optimum nor practical because they are zero at the edge; most dish designs exhibit considerable feed spillover. Special designs for very low noise antennas however might utilize zero edge illumination.⁹ In array antennas, Taylor distributions are often used as the excitation envelope. The distributions discussed in

TABLE I
CHARACTERISTICS OF 10 λ APERTURES

| | Uniform | Taylor | | |
|-----------------------------|---------|--------|-------|-------|
| | | 25 | 30 | 35 |
| Sidelobe level, decibels | -17.56 | -25 | -30 | -35 |
| \bar{n} (see ref. 8) | | 3 | 4 | 5 |
| Beamwidth, degrees | 5.85 | 6.90 | 7.25 | 7.60 |
| Edge illumination, decibels | 0 | -8.1 | -10.7 | -14.0 |

⁶ R. C. Hansen, L. L. Bailin, and R. W. Rutishauser, "On computing radiation integrals," *J. Assoc. Comput. Mach.*, vol. 2, pp. 28-31; February, 1959.

⁷ T. T. Taylor, "Design of Circular Apertures for Narrow Beamwidth and Low Sidelobes," Microwave Laboratory, Hughes Aircraft Co., Culver City, Calif., Tech. Memo. 372; August, 1954.

⁸ R. C. Hansen, "Tables of Taylor Distributions for Circular Aperture Antennas," Microwave Laboratory, Hughes Aircraft Co., Culver City, Calif., Tech. Memo. 587; February, 1959.

⁹ R. C. Hansen, "Low noise antennas," *Microwave J.*, vol. 2, pp. 19-24; June, 1959.

this section were chosen to relate closely to actual antenna excitations.

For patterns at distances which are integral multiples of λ , spherical Hankel functions were computed for $kR = 10\pi, 12\pi, 15\pi, 18\pi, 20\pi, 25\pi, 30\pi, 40\pi, 50\pi, 80\pi$, and 100π . The distances chosen are in units of $2D^2/\lambda$; with $x = R/2D^2/\lambda$, the distances are $x = \infty, 0.25, 0.125, 0.075, 0.05, 0.0375$. Figs. 2-5 depict a superposition of the field patterns at various distances for the uniform and 25-, 30-, 35-db Taylor distributions. From the axial power curves of the next section, it can be observed that this range of x covers the far field region and two oscillations in the Fresnel region (to within a diameter of the aperture). All of the curves in these figures refer to a $ka = 10\pi$ (10 λ diameter) aperture. It will be shown in Section V that this size aperture can be used to describe some fundamental characteristics of radiating surfaces. Just as in optics, and in the lens fields described by Bachynski and Bekefi,¹⁰ the pattern amplitude reaches a maximum near $D^2/5^2 \lambda$, which is slightly closer than the point where the aperture subtends a single Fresnel zone ($D^2/4 \lambda - \lambda/2$). The main beam bifurcates into a forked beam near $D^2/8 \lambda$.

Several items of interest can be deduced from these plots. At half the distance to the far field transition point ($x = 0.25$) the first sidelobe has been swallowed by the main beam. The second lobe (numbering lobes according to the far field patterns) has been raised slightly; other lobes are unchanged, except for wide angle lobes. The surprising fact noted is that at $x = 0.25$ the 60° sidelobe ratio is larger than at either $x = 0.125$ or $x = \infty$. This will be discussed further in Section V. In general, as the distance decreases the sidelobe envelope rises and the minima become shallow. This effect is especially pronounced at small observation angles. As an observer moves closer to the antenna, the pattern becomes smoother and has less "character." All these observations apply to each of the distributions. As the design sidelobe ratio increases for fixed x , small angle sidelobes are swallowed up. For 35 db and $x = 0.25$, two of the sidelobes have been devoured by the main beam. Kelly¹¹ has shown that for 50 db even at $x = 1$, one sidelobe has been swallowed. Thus the very low sidelobe designs, around 35 to 40 db, are on the borderline of approximating the desired pattern at D^2/λ . At any fixed distance the sidelobe envelope appears, from these limited data, to be the parameter most sensitive to changes in illumination. This information is pertinent to the problem of computing far field characteristics from limited Fresnel region measurements.

The next section will derive Fresnel formulas and compare the patterns obtained with these exact results.

IV. FRESNEL FORMULAS

Although Fresnel formulas are often written directly from a Kirchhoff integral or similar scalar integral, it will be more precise and more illuminating to start directly with (9), discarding all terms of order greater than $(1/r)$. (See the Appendix for a definition of the Fresnel field.) The result, again writing only the θ component, is

$$E_\theta = \frac{-j\eta \cos \theta \cos \phi}{4\pi} \int_0^{2\pi} \int_0^{ka} \frac{e^{jkR}}{kr} f(k\rho) k\rho dk\rho d\beta. \quad (15)$$

The Fresnel approximation assumes that the kr in the denominator is constant and equal to kR , and retains only the linear and the quadratic terms (in ρ/R) in the binomial expansion of the phase factor. Thus

$$\begin{aligned} r &= [R^2 + \rho^2 - 2R\rho \sin \theta \cos(\phi - \beta)]^{1/2} \\ &= R - \rho \sin \theta \cos(\phi - \beta) \\ &\quad + \frac{\rho^2}{2R} [1 - \sin^2 \theta \cos^2(\phi - \beta)] + O\left(\frac{\rho^3}{R^2}\right) \end{aligned} \quad (16)$$

and the field component is

$$E_\theta = \frac{-j\eta e^{jkR} \cos \theta \cos \phi}{4\pi kR} \cdot I. \quad (17)$$

A straightforward change of variable (for uniform illumination) gives

$$\begin{aligned} I &= \int_0^{2\pi} \int_0^{ka} \exp \left[jk\rho \sin \theta \cos \beta \right. \\ &\quad \left. - j \frac{k\rho^2}{2R} (1 - \sin^2 \theta \cos^2 \beta) \right] k\rho dk\rho d\beta. \end{aligned} \quad (18)$$

To perform the β integration, the integral is rearranged and the quadratic phase term is expanded in series. Then the β integral is

$$\begin{aligned} \exp \left(\frac{-jk\rho^2 \cos^2 \theta}{2R} \right) \sum_{N=0}^{\infty} \left(\frac{-jk\rho^2}{2R} \right)^N \frac{\sin^{2N} \theta}{N!} \int_0^{2\pi} \sin^{2N} \beta \\ \cdot \exp(jk\rho \sin \theta \cos \beta) d\beta \end{aligned} \quad (19)$$

which with recourse to tables¹² gives

$$\begin{aligned} I &= 2\sqrt{\pi} \sum_{N=0}^{\infty} \frac{(-j)^N \Gamma(N + \frac{1}{2}) \sin^N \theta}{N! (kR)^N} \int_0^{ka} (k\rho)^{N+1} \\ &\quad \cdot \exp \left(\frac{-jk\rho^2 \cos^2 \theta}{2R} \right) J_N(k\rho \sin \theta) dk\rho. \end{aligned} \quad (20)$$

From Watson¹³ or Walker,¹⁴ the integral in (20) can be expressed as a pair of Lommel functions of two variables, and the field becomes

¹² W. Grobner and N. Hofreiter, "Integraltafel," Springer-Verlag, Wien, Austria, pt. 2, 334.58a, 1958.

¹³ G. N. Watson, "Bessel Functions," Oxford University Press, New York, N. Y., p. 541; 1952.

¹⁴ J. Walker, "The Analytical Theory of Light," Cambridge University Press, Cambridge, Eng., pp. 130, 400; 1904.

¹⁰ M. P. Bachynski and G. Bekefi, "Aberrations in circularly symmetric microwave lenses," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-4, pp. 412-421; July, 1956.

¹¹ K. C. Kelly, Hughes Aircraft Co., Culver City, Calif., unpublished paper.

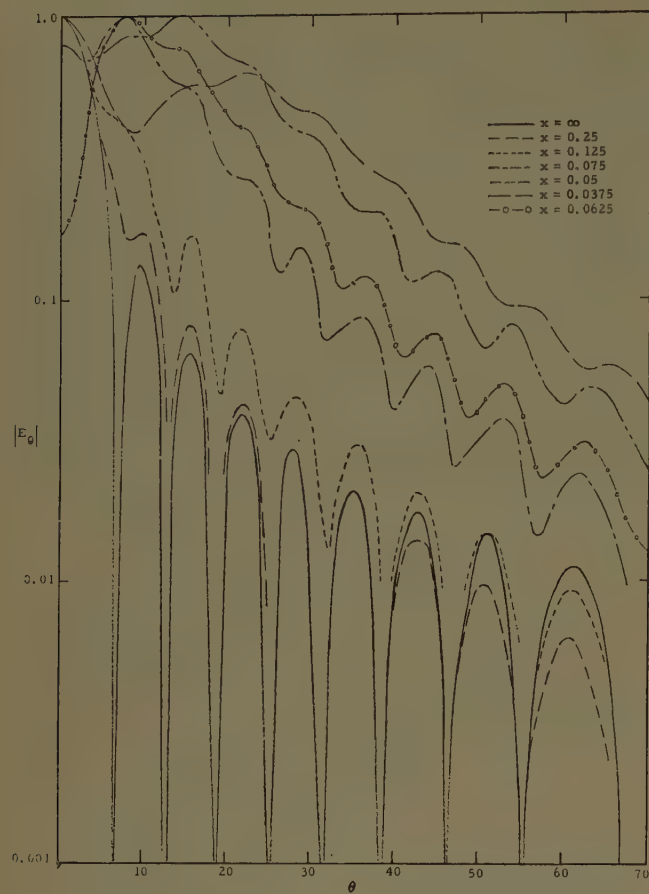


Fig. 2—Elevation patterns—uniform illumination.

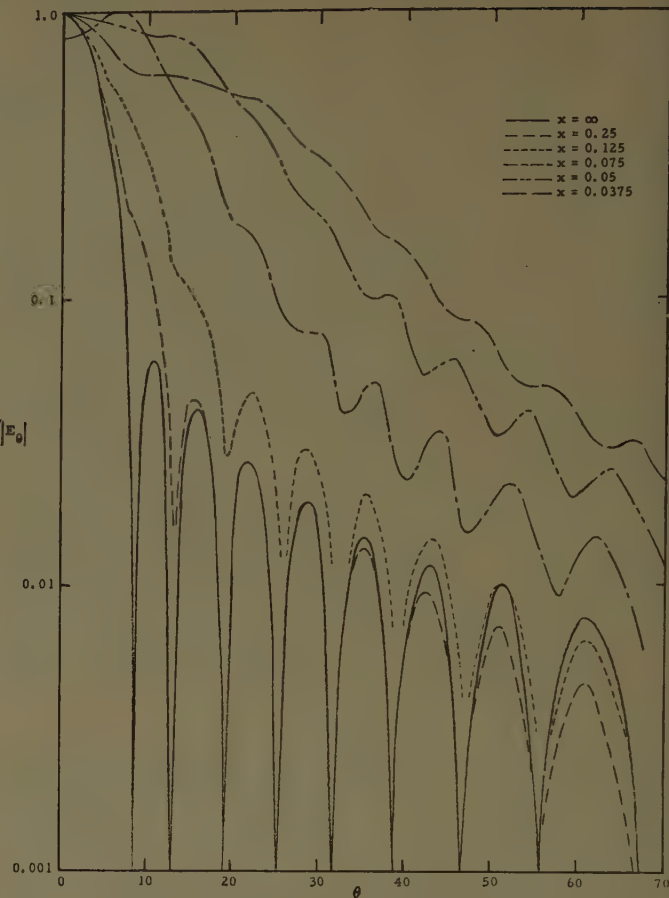


Fig. 3—Elevation patterns—25-db Taylor.

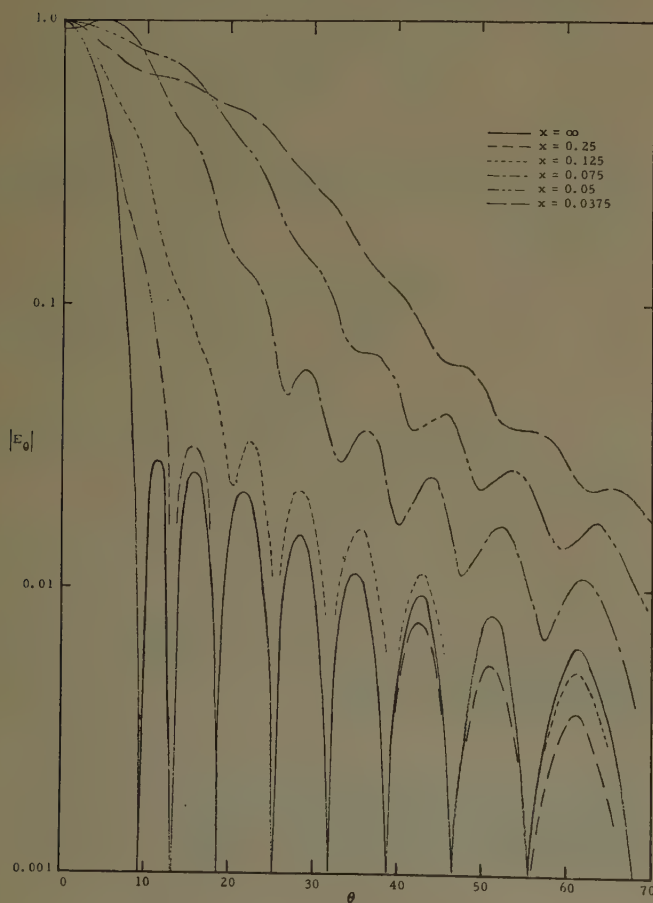


Fig. 4—Elevation patterns—30-db Taylor.

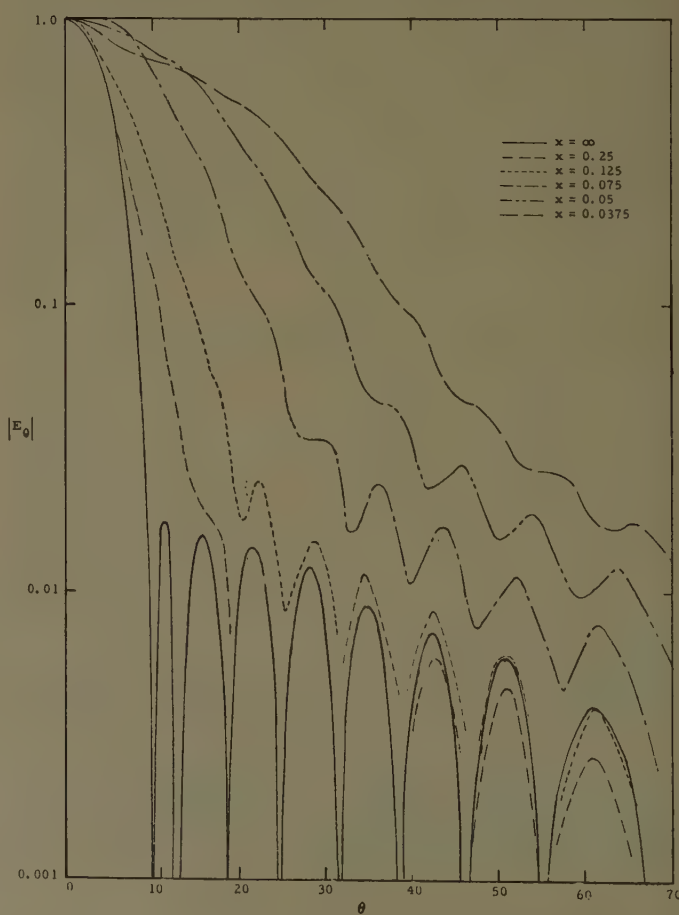


Fig. 5—Elevation patterns—35-db Taylor.

$$E_\theta = \frac{-j\eta \cos \theta \cos \phi}{2\sqrt{\pi}} \sum_{N=0}^{\infty} \frac{(-j)^N \Gamma(N + \frac{1}{2}) \sin^N \theta u^N \exp\left(\frac{-jw}{2}\right)}{N!(kR)^N w^{N+1}} \cdot [U_{N+1}(w, u) + jU_{N+2}(w, u)] \quad (21)$$

where

$$w = \frac{ka^2}{2R} \cos^2 \theta \quad \text{and} \quad u = ka \sin \theta.$$

The small angle Fresnel formula is obtained either by discarding the $\sin^2 \theta$ term in (18) or by discarding $\sin^N \theta$ factor in the above equation, where $N > 0$. This gives the first term answer of Hu.¹⁵ Because Lommel functions are very imperfectly tabulated (existing tables are those of Lommel and Hu), calculations are usually made in terms of a series of Bessel functions as

$$U_N(w, u) = \sum_{\rho=0}^{\infty} \frac{(-1)^\rho w^{N+2\rho}}{u^{N+2\rho}} J_{N+2\rho}(u). \quad (22)$$

This result can be inserted into (21) to give E_θ in terms of Bessel functions. The small angle result, the first term in (21), is therefore

$$E_\theta \simeq \frac{-j\eta}{2\gamma} \cos \theta \cos \phi \exp\left(\frac{j\gamma}{2}\right) [U_1(\gamma, u) + jU_2(\gamma, u)] \quad (23)$$

where w has been replaced by

$$\gamma = \frac{ka^2}{2R}.$$

This small angle result for uniform illumination is the same as that obtained by Lommel¹⁶ in 1884 for the problem of diffraction through a circular hole in a screen. His results were recapitulated by Walker¹⁴ in 1902. The series of Bessel functions, called Lommel functions of two variables, were studied extensively by Lommel and Walker. An excellent derivation of the field with $(1 - k^2 \rho^2)^N$ illumination taper is given by Hu.¹⁵ Hu obtains solutions in a Bessel series where the $N=0$ case yields the results described above.

For the Fraunhofer field, w is small and the Lommel functions can be approximated by the first term in (22). The result is then

$$E_\theta = -j\eta \cos \theta \cos \phi \frac{J_1(u)}{2u}. \quad (24)$$

This equation is usually obtained by a saddle-point integration of the scalar Kirchhoff integral. The results of calculations using the small angle (23) and general (21)

Fresnel formulas are plotted in Figs. 6-9 for $x=0.25, 0.125, 0.0625, 0.05$. Curves from the exact formulas are also shown. It is apparent, as expected, that the two Fresnel results diverge for large angles. At $x=0.25$ and 0.125 the general Fresnel results are excellent except in the depth of minima, whereas the small angle results are too large, both at the maxima and minima. At the two distances closer to the aperture, both Fresnel results display extrema that are shifted. At $x=0.05$, the Fresnel results are zero on axis; this represents the greatest error as the actual field is finite on axis. It is apparent from the poor correspondences of Fresnel and exact results in Figs. 8 and 9 that for the 10λ aperture, the Fresnel results are inaccurate for $x < 0.0625$, corresponding to the first minima. The lower limit established by Polk¹⁷ for a square aperture (phase error of $\lambda/8$ at edge) gives

$$R \geq \frac{\sqrt[3]{2}}{2} D \left(\frac{D}{\lambda}\right)^{1/3}, \quad (25)$$

which for the case considered in this paper is $x \geq 0.068$. Hu's lower limit,¹⁵ twice (25), is conservative for this example. The question of the lower limit of the Fresnel region for very large apertures remains to be answered. For the 10λ example, the lower limit is 12.5λ , or slightly more than one aperture diameter. Eq. (25) indicates that larger apertures have Fresnel region lower limits equal to several aperture diameters.

V. AXIAL POWER DENSITY

Of appreciable interest is the variation of power density along the disc axis. This information is pertinent to personnel safety problems and to power transmission systems since the latter will probably use antennas located within each other's Fresnel regions. For the circular aperture, two simple cases have been computed from the Fresnel formulas for apertures with uniform and $(1 - k^2 \rho^2)$ tapered illumination. The formulas are

$$P = \frac{16.2}{x^2} \left| \int_0^1 \exp \left[\frac{j\pi}{8x} (1 - \xi^2) \right] \xi d\xi \right|^2 = 13.14 \left[1 - \cos \frac{\pi}{8x} \right] \quad (26)$$

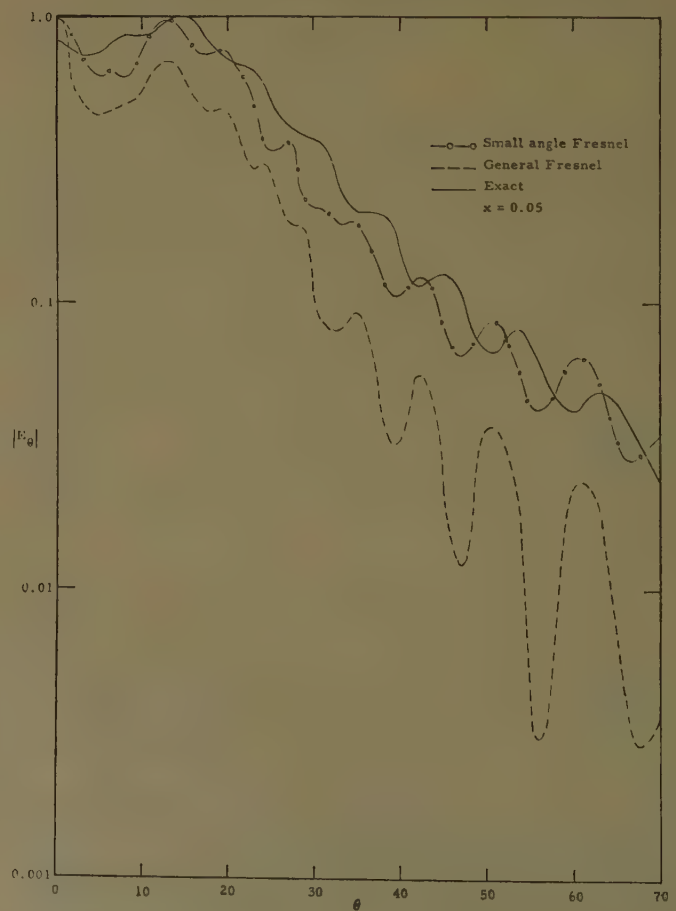
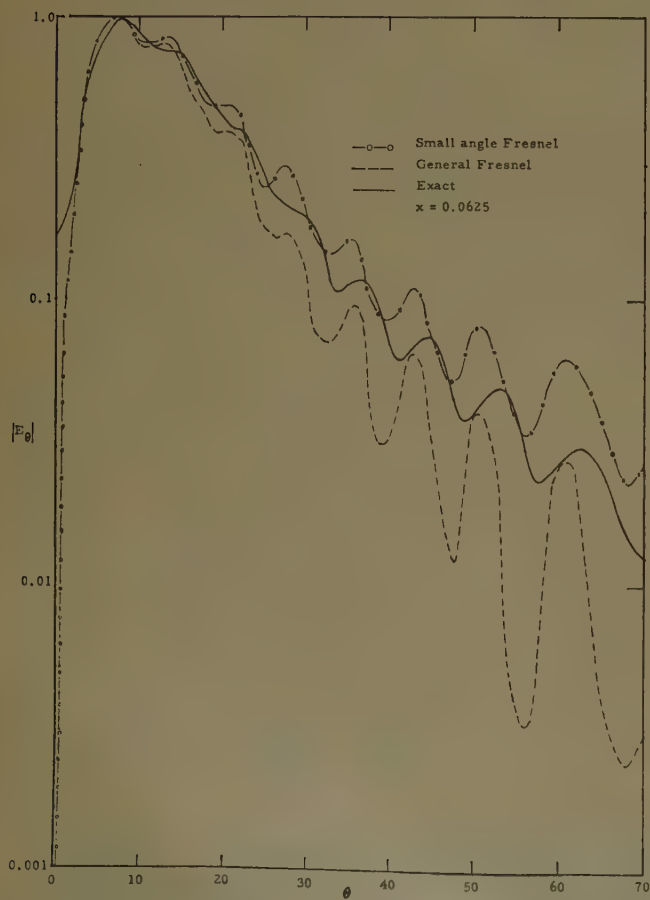
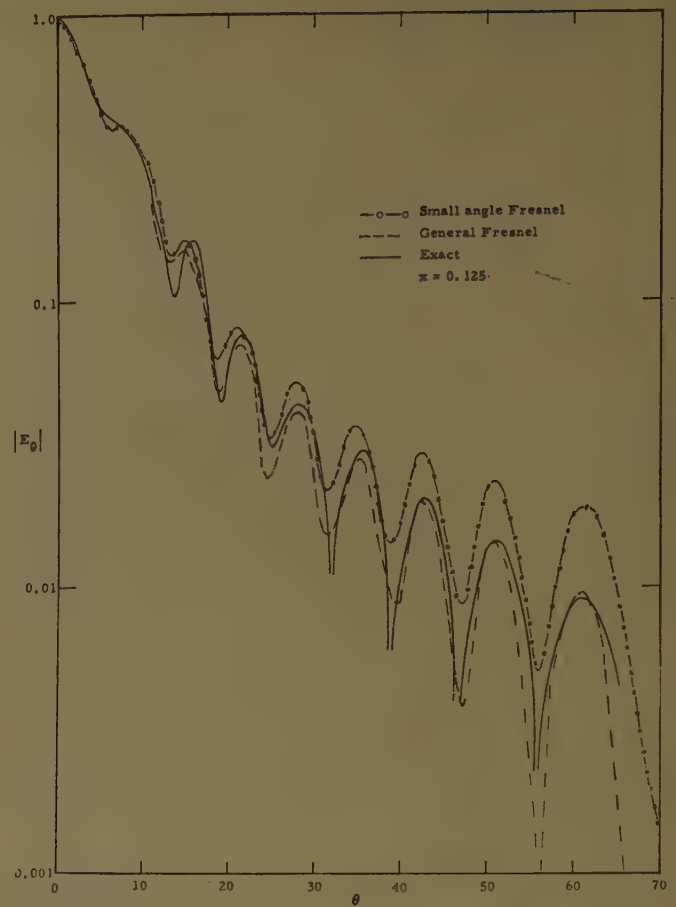
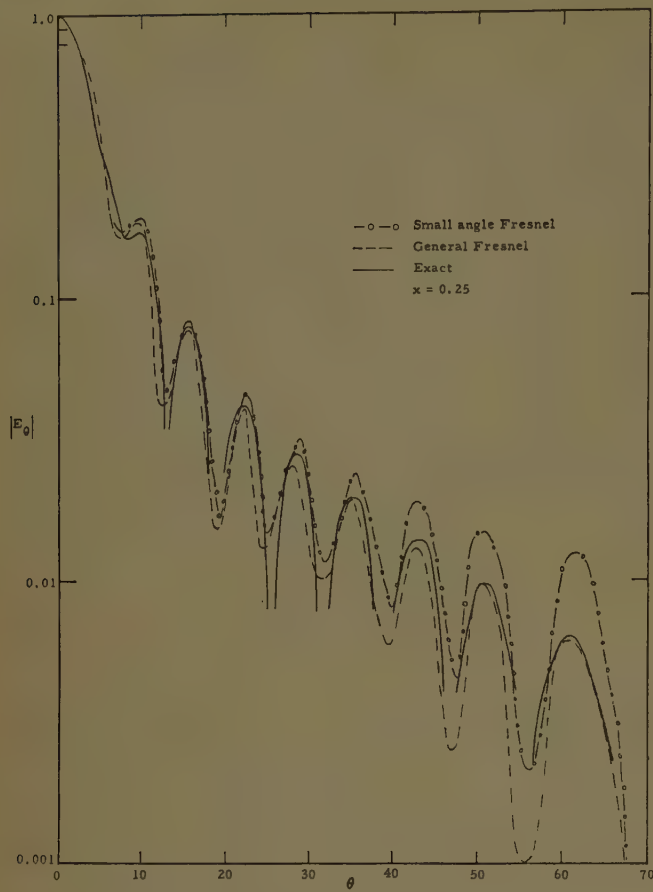
for uniform illumination, and

$$P = \frac{16.1}{x^2} \left| \int_0^1 (1 - \xi^2) \exp \left(\frac{j\pi}{8x} (1 - \xi^2) \right) \xi d\xi \right|^2 = 26.1 \left[1 - \frac{16x}{\pi} \sin \frac{\pi}{8x} + \frac{128x^2}{\pi^2} \left(1 - \cos \frac{\pi}{8x} \right) \right] \quad (27)$$

¹⁵ M. K. Hu, "Study of Near-Zone Fields of Large Aperture Antennas," Res. Inst., Syracuse University, Syracuse, N. Y., Final Rept., Part 2, RADC-TR-57-126B; April, 1957.

¹⁶ E. Lommel, "Theoretical and experimental investigations of diffraction phenomena at a circular aperture and obstacle," *Bayerisch. Akad. d. Wiss.*, vol. 15, p. 233; 1884. Translated by G. Bekefi and G. A. Woonton, McGill University, Montreal, Canada.

¹⁷ C. Polk, "Optical Fresnel-zone gain of a rectangular aperture," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-4, pp. 65-69; January, 1956. Also Ph.D. dissertation, University of Pennsylvania, Philadelphia, Pa.; 1956.



for the $(1 - k^2 p^2)$ tapered illumination. In both cases (see Figs. 10 and 11) the constants have been chosen to make P unity at $x = 1$; and the distance is given in terms of

$$x = \frac{R}{2D^2/\lambda}.$$

Curves based on the exact field calculations are also included. The Fresnel zone phenomenon is displayed by the alternate addition and cancellation of zones in Fig. 10. In the next figure, the effect of the zones is "smeared out" by the incomplete interference caused by the taper. The phenomenon of complete addition and cancellation is unique to the uniformly illuminated circular aperture. The uniform square aperture shows the characteristic smearing of all tapered illuminations, due to the corners. It may be noted that the exact and Fresnel result agree very well for $x > 0.1$. Note that the power varies with distance as $1/R^2$ up to D^2/λ ; this approximate behavior will prove to be valid in all subsequent power density calculations. Thus D^2/λ is an excellent demarcation line between far and Fresnel fields, as indicated in the Appendix. In Fig. 11 the 25-db curve exhibits a first maximum that is smaller in amplitude than the subsequent maximum. In Fig. 12 power density curves are plotted for the 30- and 35-db Taylor distributions. Figs. 10-12 show that the shape of the aperture distribution, particularly the height of the pedestal, affects the axial power curve appreciably.

Fig. 13 shows, for the uniform illumination case, the variation of the amplitude of one of the wide angle (60°) sidelobes with x . The amplitude increases monotonically in the range shown, but the increase is slower than the increase of amplitude on axis (see Fig. 10). The sidelobe ratio is consequently slightly higher than the far field value, with a maximum near $x = 0.3$. For points closer to the aperture, the sidelobe amplitude increases quite rapidly, lowering the sidelobe ratio.

In (9), only one term is $1/r$ in order, and this term is the genesis of Fresnel and Fraunhofer approximations. The series resulting from this term contains the $A_1(N)B_1(N)$ terms in (14). In order to establish importance of the higher order terms, the $1/r$ points have been plotted in Figs. 10-12. The difference between these points and the total curve is the higher order contribution, which includes the edge correction term. The edge integral has, in general, a very minor effect on the axial distribution. For wide angles the edge contributions are not negligible, but near the axis they are very small except very close to the aperture. This substantiates the work of Chang and Silver,¹⁸ and Plonsey,¹⁹ who show that the edge integral contribution is a small perturbation.

Fig. 14 shows theoretical and experimental curves after Sterns.²⁰ Sterns assumed a tapered illumination with a small pedestal; his curves like those calculated from the exact formulas show the first peak smaller than peaks closer in to the aperture. The experimental curve fits the 25-db Taylor curve of Fig. 11 fairly well.

With reference to the axial power curves, the $ka = 10\pi$ aperture includes the principal maximum ($x \approx 0.1$), one valley, and a second maximum; $x = 0.05$ represents a distance one diameter away from the aperture. A larger diameter aperture would include more oscillations, but would not contribute significant knowledge. The oscillations damp out in all but the uniform case. So it is felt that the examples treated here contain the essence of the problem.

VI. CONCLUSION

The general behavior of the field with distance shows a progressive deterioration in the sidelobe region with decreasing distance. At $R = D^2/2\lambda$, the first sidelobe (counting lobes in the far field pattern) has been swallowed by the main beam skirt although the principal part of the main beam is unchanged as are the remaining sidelobes. As distance decreases, the main beam broadens and continues to absorb sidelobes, and the level of remaining sidelobes is raised. For very low sidelobe designs, around 35 to 40 db, a desired far field pattern can be only loosely approximated at distance $R = D^2/\lambda$.

From very large distances to $R = D^2/\lambda$, the axial field intensity is close to a $1/R$ variation, for all the aperture distributions. Thus, D^2/λ is a good transition point between far and Fresnel regions. Practical aperture distributions exhibit a first maximum in axial field vs distance (at about $D^2/5\lambda$) that is smaller than subsequent maxima in contrast to the behavior of classical aperture distributions. This effect appears to be dependent upon the illumination pedestal, but not through the edge integral terms. In fact the latter have a small effect, and do not even affect the curve shapes close to the antenna.

For distances greater than about $D^2/5\lambda$, both the axial power density and the wide angle sidelobe level are monotonic increasing functions with decreasing R . However the axial density increases more rapidly around $D^2/3\lambda$ producing the interesting result that the sidelobe ratio for wide angle lobes actually increases over the far field value.

The Fresnel formulas are useful from D^2/λ to at least the first minimum in axial field. Small angle results are, as might be expected, poor for the wide angle sidelobes, giving maximum and minimum values that are too large. For the 10λ aperture studied, the lower Fresnel limit is close to $D^2/7\lambda$. It appears for larger apertures that this limit would be for R equal to several aperture diameters.

¹⁸ W. S. Chang and S. Silver, "Current Distributions on Circular Cylindrical Reflectors," University of California, Berkeley, ERL Rept. 60-193; August, 1957.

¹⁹ R. Plonsey, "Diffraction by Cylindrical Reflectors. II," University of California, Berkeley, ERL Rept. 60-183; July, 1957.

²⁰ W. G. Sterns, "Near-Zone Field Studies of Quasi-Optical Antennas," M.S. thesis, University of California, Berkeley; circa 1954.

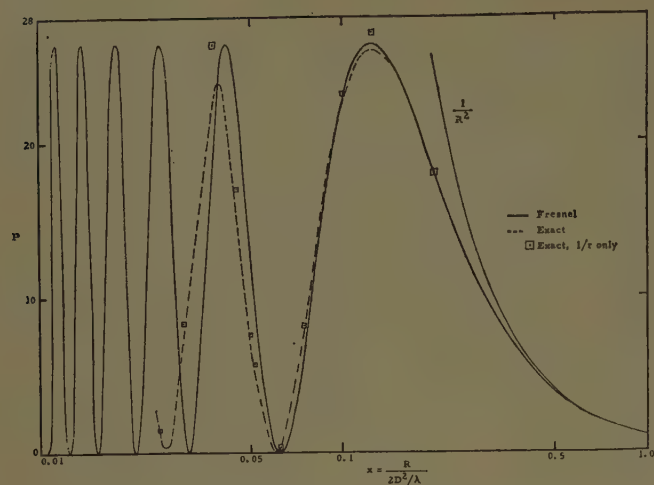


Fig. 10—Axial power density—uniform illumination.

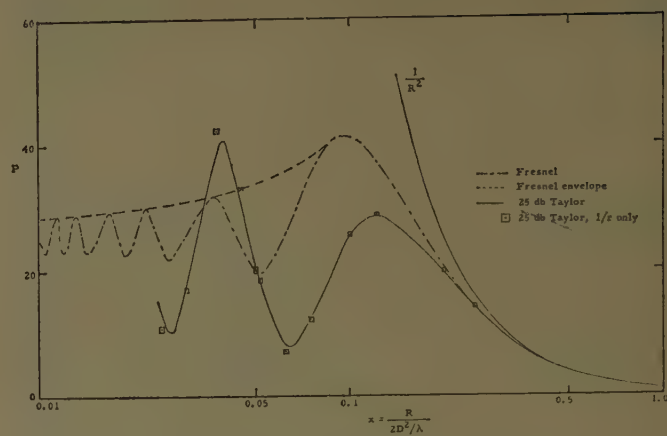


Fig. 11—Axial power density—tapered illumination.

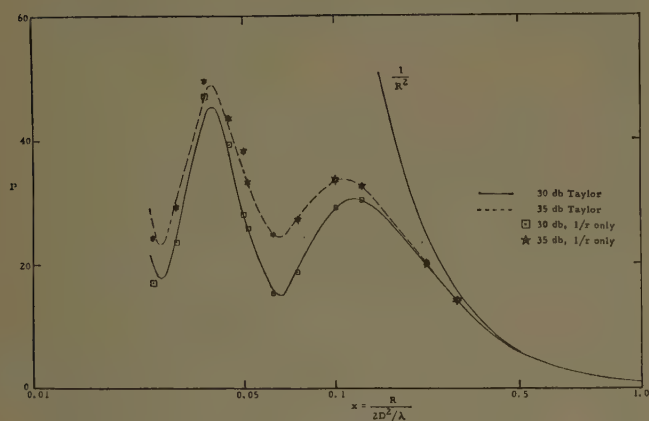


Fig. 12—Axial power density—Taylor apertures.

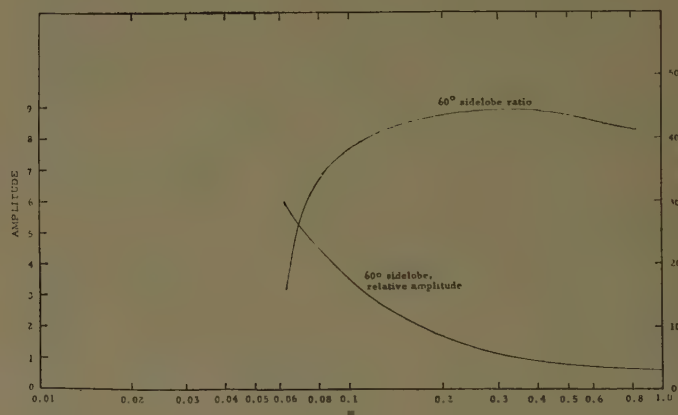


Fig. 13—Wide angle sidelobes.

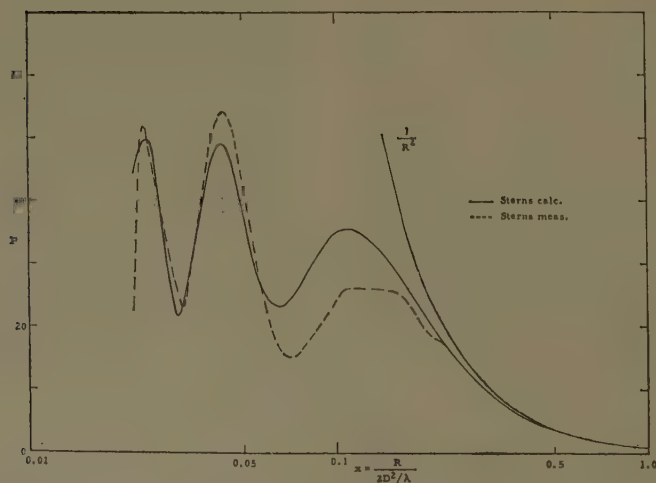


Fig. 14—Experimental axial power density.

An excellent approximation to the exact field is obtained by using only the $1/r$ term in the radiation integral; this term contains the Fresnel result.

APPENDIX

DEFINITIONS OF FIELD TERMS

Far Field Region—That region in which the field intensity varies as $1/R$, where R is the distance to an origin in the local neighborhood of the antenna. The approximation for the field of each element (Huygen's source) of the aperture consists of a constant amplitude factor and a linear (first order) phase factor. If the antenna has a well defined aperture D in a given aspect, the far region in that aspect is commonly taken to extend from a distance D^2/λ from the aperture to infinity.

Fraunhofer Region—Same as far field region.

Fresnel Region—That region of the field wherein the approximation of the field of each element (Huygen's source) of the aperture consists of a constant amplitude factor and two phase factors, a linear and a quadratic

(second order). The Fresnel field is a $1/r$ field in aperture coordinates, but requires many terms of the form $1/R^n$ in space coordinates. The Fresnel region extends from a distance of several times D from the aperture to D^2/λ .

Near Field Region—That region from the antenna surface to the Fresnel region, usually several diameters away from the surface. The near field requires many terms of the form $1/r^n$ both in aperture and space coordinates. All space is divided into near, Fresnel and Fraunhofer regions.

ACKNOWLEDGMENT

The authors wish to express thanks to Prof. S. Silver, and Dr. R. W. Bickmore for their suggestions, to Miss J. Van Hooven for a superb computer programming job, to Mrs. H. Arens and Miss C. Hasson for diligent evaluation of approximate formulas, and finally to colleagues in the Antenna Research Department, Hughes Aircraft Company for their aid and encouragement.



Back Scattering at High Frequencies from a Conducting Cylinder with Dielectric Sleeve*

RALPH D. KODIS†

Summary—The computational difficulties connected with the problem of high-frequency back scattering from a conducting cylinder with dielectric sleeve arise from the slow convergence of the conventional Fourier representation of the field and are compounded by the involved nature of its Fourier coefficients. In addition, the numerical results give no physical insight into the complicated structure of the back-scattering functions. An alternative representation of the diffracted amplitude as a series of radial eigenfunctions has the advantage of rapid convergence at high frequencies but presents difficulties of its own since one must find not only the complex coefficients of the expansion but also the complex indices for which the coefficients are to be evaluated. Some of these difficulties can be avoided by transforming the radial representation into a sum of terms, one of which is a well-known form of the diffracted amplitude from a conducting cylinder whose radius is the same as the outer radius of the dielectric sleeve. The second term, which contains the effect of the sleeve, turns out conveniently to be an infinite integral over a real variable. An expansion of its integrand leads to a series of terms which are analogous to optical rays. When the over-all cylinder radius is large, each of these terms has a stationary phase approximant over a certain range of dielectric thickness and relative dielectric constant. Only over this range does the corresponding ray contribute to the back-scattered amplitude. The detailed evaluation of three of these integrals gives results which account for some of the features of the back-scattering functions.

It is well established by now that the representation of wave solutions for separable geometries in terms of suitable radial eigenfunctions has certain advantages for high-frequency calculations. A general method for finding such representations has been developed and discussed by Titchmarsh¹ and Marcuvitz,² among others. We shall begin by outlining the application of this method to the task of determining the wave solution outside an infinitely long, perfectly conducting cylinder with dielectric sleeve when the cylinder is excited by an external harmonic ($e^{-i\omega t}$) electric line source parallel to its axis. The symmetry of this configuration reduces the problem to scalar diffraction in a plane with the geometry shown in Fig. 1. The formal solution is an appropriate representation of the Green's function which satisfies the following system of differential equations and boundary conditions:

1) Wave equations.

$$(\nabla^2 + k_0^2)G_0(r, r') = -\delta(r - r'); \quad r, r' \geq b. \quad (1a)$$

$$(\nabla^2 + k^2)G_1(r, r') = 0; \quad a \leq r \leq b, \quad r' \geq b. \quad (1b)$$

* This research was supported jointly by the Electronics Research Directorate of the AF Cambridge Research Center under Contract AF 19(604)-4561, and by the Office of Naval Research under contract Nonr-562(24).

† Div. of Engineering, Brown University, Providence, R. I.

¹ E. C. Titchmarsh, "Eigenfunction Expansions," Oxford University Press, New York, N. Y., 1959.

² N. Marcuvitz, *Commun. Pure Appl. Math.*, vol. 4, pp. 284-293; August, 1951.

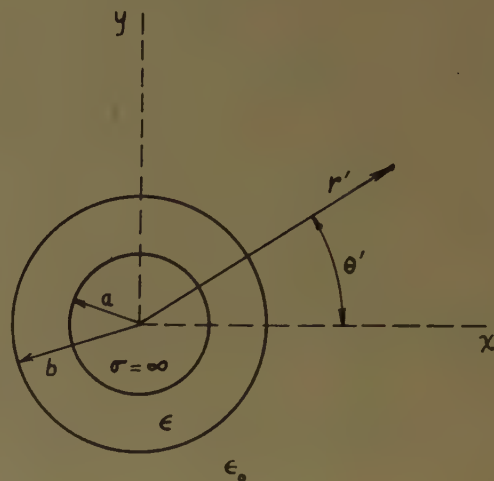


Fig. 1.

2) Boundary conditions.

$$G_1(a, r') = 0 \quad [E_r = 0 \text{ at } r = a]. \quad (2a)$$

$$G_1(b, r') = G_0(b, r') \quad [E_z \text{ is continuous at } r = b]. \quad (2b)$$

$$\partial_r G_1(b, r') = \partial_r G_0(b, r') \quad [B_\theta \text{ is continuous at } r = b]. \quad (2c)$$

3) The radiation condition.

In (1), k is the wave number ω/v , where v is the phase velocity characteristic of the medium. The coplanar position vectors \mathbf{r}' and \mathbf{r} locate the source and observation points relative to an origin on the cylinder axis. The vectors \mathbf{a} and \mathbf{b} of (2) terminate on the inner and outer circumference of the sleeve, respectively.

In order to determine the Green's function, G_0 , associated with the above problem, we require characteristic functions which are solutions of a set of three independent ordinary differential equations plus boundary conditions corresponding to (2). These differential equations can be obtained from (1) by separation of the polar variables (r, θ) . Denoting the separation parameter by the symbol λ , we find that in $r, r' \geq b$ the characteristic functions are

$$g(\theta, \theta', \lambda) = \frac{i}{2\nu} e^{i\nu|\theta - \theta'|}; \quad (\nu = \sqrt{\lambda}, \quad 0 < \arg \nu < \pi) \quad (3)$$

$$f_0(r, r', \lambda) = \frac{\pi i}{4} \left[H_{\nu}^{(2)}(k_0 r_<) - \frac{H_{\nu}^{(2)}(\beta_0) X_{\nu}^{(2)}}{H_{\nu}^{(1)}(\beta_0) X_{\nu}^{(1)}} H_{\nu}^{(1)}(k_0 r_<) \right] H_{\nu}^{(1)}(k_0 r_>), \quad (4)$$

where $H_{\nu}^{(1)}$, $H_{\nu}^{(2)}$ are Hankel functions of the first and second kind, and $r_>$, $r_<$ signify the larger and the smaller

of the variables r and r' . In addition, these expressions contain the following shorthand notations:

$$\alpha = ka, \quad \beta = kb, \quad \beta_0 = k_0b; \quad (5)$$

$$C_\nu(kr) = H_\nu^{(1)}(\alpha)H_\nu^{(2)}(kr) - H_\nu^{(2)}(\alpha)H_\nu^{(1)}(kr); \quad (6)$$

$$X_\nu^{(1,2)} = \beta \frac{C_\nu'(\beta)}{C_\nu(\beta)} - \beta_0 \frac{H_\nu^{(1,2)'}(\beta_0)}{H_\nu^{(1,2)}(\beta_0)}. \quad (7)$$

In accordance with the formula,

$$G_0(r, r') = \frac{1}{2\pi i} \oint d\lambda g(\theta, \theta', \lambda) f_0(r, r', \lambda), \quad (8)$$

the convolution of (3) and (4) in the complex λ plane leads to the required exterior Green's function. For our purposes the integration contour encloses the branch cut of $g(\theta, \theta', \lambda)$ which is taken along the real axis of the λ plane. Eliminating the function $X_\nu^{(2)}$ in favor of $X_\nu^{(1)}$ and changing variables from λ to ν , we obtain

$$G_0(r, r') = \frac{i}{8} \int_{-\infty}^{\infty} d\nu [H_\nu^{(1)}(\beta_0)H_\nu^{(2)}(k_0r_<) - H_\nu^{(2)}(\beta_0)H_\nu^{(1)}(k_0r_<)] \frac{H_\nu^{(1)}(k_0r_>)}{H_\nu^{(1)}(\beta_0)} e^{i\nu|\theta-\theta'|} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{d\nu}{X_\nu^{(1)}} \frac{H_\nu^{(1)}(k_0r_<)H_\nu^{(1)}(k_0r_>)}{H_\nu^{(1)}(\beta_0)H_\nu^{(1)}(\beta_0)} e^{i\nu|\theta-\theta'|}. \quad (9)$$

It is easy to show that the first integral of (9) is the Green's function for a perfectly conducting cylinder of radius b with Dirichlet boundary condition. The effect of the sleeve on this field resides in the second integral since both the thickness and the refractive index of the dielectric are parameters only of the function $X_\nu^{(1)}$.

The expression given in (9) represents the total field at any point outside the cylinder in the most general situation, that is, when the excitation is due to a line source at a finite distance. In order to specialize this result to the back-scattered amplitude with plane wave excitation, we set $\theta = \theta'$ and let $r_> \rightarrow \infty$. Furthermore, since our discussion will be restricted to the far-zone field, we shall be interested in the region where $r_< = r$ is large but finite. The procedure involves first the substitution of the leading term of the Debye expansion in the combination $H_\nu^{(1)}(k_0r_<)H_\nu^{(1)}(k_0r_>)$, then the normalization of the source function in such a way that the incident plane wave has unit amplitude. Thus, the far-zone amplitude in the direction of the source is found to have the asymptotic form,

$$\psi(r) \sim \psi_b(r) + \frac{2}{\pi i} \left(\frac{2}{\pi k_0 r} \right)^{1/2} \cdot \exp i \left(k_0 r - \frac{\pi}{4} \right) \int_{-\infty}^{\infty} \frac{d\nu}{X_\nu^{(1)} [H_\nu^{(1)}(\beta_0)]^2} e^{-i\nu\pi}, \quad (10)$$

where $\psi_b(r)$ is the amplitude (incident + scattered) associated with a perfectly conducting cylinder of radius b .

In principle, $\psi_b(r)$ is a known function. Its form at high frequencies may be determined from the first inte-

gral of (9) or from straightforward considerations of ray optics. Both calculations lead to the result,

$$\psi_b(r) \sim e^{-ik_0 r} - \left(\frac{\beta_0}{2\pi k_0 r} \right)^{1/2} \exp i(k_0 r - 2\beta_0), \quad (11)$$

where the two terms represent the incident plane wave and the far-zone back-scattered amplitude, respectively. What remains to be determined is how this field is modified by the dielectric properties of the outer sleeve of the cylinder. To this end we proceed to consider the integral,

$$\int_{-\infty}^{\infty} I(\nu) d\nu = \int_{-\infty}^{\infty} \frac{d\nu}{X_\nu^{(1)} [H_\nu^{(1)}(\beta_0)]^2} e^{-i\nu\pi}. \quad (12)$$

The asymptotic evaluation of (12) to be outlined below makes use of Hankel function approximations which are valid in the frequency region $ka \gg 1$ and over the interval $\nu^2 < \alpha^2$. In this interval, we may expect to find points of stationary phase where the leading contributions to the integral are made. The exponential behavior of the Hankel function outside the interval will insure that the contribution there is of lower order. Thus, the appropriate Debye expansions for functions of argument β_0 are

$$H_\nu^{(1)}(\beta_0) \sim 2^{1/2} e^{-i(\pi/4)} (\pi\beta_0 \sin \eta_0)^{-1/2} \exp i\beta_0 f(\eta_0), \quad (13a)$$

$$H_\nu^{(1)'}(\beta_0) \sim i \sin \eta_0 H_\nu^{(1)}(\beta_0), \quad (13b)$$

where

$$\cos \eta_0 = \frac{\nu}{\beta_0}; \quad f(\eta_0) = \sin \eta_0 - \eta_0 \cos \eta_0. \quad (14)$$

With similar approximations for $H_\nu^{(1,2)}(\alpha)$ and $H_\nu^{(1,2)}(\beta)$, the integrand of (12) takes the form,

$$I(\nu) \sim -\frac{\pi}{2} \left[\frac{\rho}{1+\rho} \frac{1-e^{2iF(\nu)}}{1+\tau e^{2iF(\nu)}} \right] e^{-2i\beta_0 f(\eta_0)-i\nu\pi}, \quad (15)$$

where

$$F(\nu) = \beta f(\eta) - \alpha f(\zeta),$$

$$\cos \zeta = \frac{\nu}{\alpha}; \quad \cos \eta = \frac{\nu}{\beta} \quad (16)$$

and

$$\rho = \frac{\beta_0 \sin \eta_0}{\beta \sin \eta}; \quad \tau = \frac{1-\rho}{1+\rho}. \quad (17)$$

Since $\beta > \beta_0$, we note that $0 < \rho < 1$; it follows that $\tau < 1$. The denominator of (15) can, therefore, be expanded in a Taylor series about the point $\tau = 0$ (i.e., $k = k_0$). After a little algebra, it is found that

$$I(\nu) \sim -\frac{\pi}{2} \left[\frac{\rho}{1+\rho} - \frac{2\rho}{(1+\rho)^2} e^{2iF(\nu)} + \frac{2\rho(1-\rho)}{(1+\rho)^3} e^{4iF(\nu)} - \dots \right] e^{-2i\beta_0 f(\eta_0)-i\nu\pi}. \quad (18)$$

A noteworthy feature of (15) and (18) is that the bracketed terms have precisely the form found for the solution of the problem of diffraction by a plane dielectric slab over a conducting interface. In the analogy, ρ and τ are related to the reflection and transmission coefficients at the dielectric-air surface. In the present context, however, they are not constants but functions of the integration variable, ν .

The remaining task is to obtain the stationary phase approximant for each integral of (18). Since the calculations become increasingly more involved for successive terms of the series, only two will be discussed in detail. The first of these is

$$K_1 = -\frac{\pi}{2} \int_{-\infty}^{\infty} \frac{\rho}{1+\rho} \exp[-2i\beta_0 f(\eta_0) - i\nu\pi] d\nu \quad (19)$$

which has its point of stationary phase at $\nu_1=0$. The Hankel function approximations are valid in this neighborhood, and the asymptotic value of K_1 may be found in the usual way. The result has the same phase as the scattered wave of (11) and may be combined with it in accordance with (10). We find in this way that the first contribution to the total back-scattered amplitude is

$$\psi_1^{\text{sc.}} = -\frac{n-1}{n+1} \left(\frac{\beta_0}{2k_0 r} \right)^{1/2} e^{ik_0 r - 2i\phi_0}, \quad (20)$$

where n is the refractive index, $k/k_0 = (\epsilon/\epsilon_0)^{1/2}$. This contribution is readily interpreted as arising from the reflection at the dielectric interface of the infinitesimal bundle of normal rays. Specifically, the ratio $(n-1)/(n+1)$ is just the reflection coefficient at the dielectric surface for normal incidence; the phase factor is proportional to the optical path length on the ray axis; and the remaining amplitude factor accounts for the spreading of the bundle.

The second term associated with (18) is

$$K_2 = \pi \int_{-\infty}^{\infty} \frac{\rho}{(1+\rho)^2} \exp i[2F(\nu) - 2\beta_0 f(\eta_0) - i\nu\pi] d\nu. \quad (21)$$

Its stationary phase point is determined by the relation

$$\nu_2^2 = \beta_0^2 - \frac{\alpha^2}{4n^2} \left[(n^2 - 1) - \frac{\beta^2}{\alpha^2} \right], \quad (22)$$

which has real solutions if, and only if, the right-hand side is positive. When this condition is satisfied, it follows from (16) and (17) that

$$(\sin \eta_0)_2 = \frac{1}{2} \frac{\alpha}{\beta} \left(n^2 - \frac{\beta^2}{\alpha^2} - 1 \right), \quad (23a)$$

$$(\sin \eta)_2 = \frac{1}{2n} \frac{\alpha}{\beta} \left(n^2 + \frac{\beta^2}{\alpha^2} - 1 \right), \quad (23b)$$

$$(\sin \zeta)_2 = \frac{1}{2n} \left(n^2 - \frac{\beta^2}{\alpha^2} + 1 \right); \quad (23c)$$

and

$$\rho_2 = \frac{(n^2 - 1) - \frac{\beta^2}{\alpha^2}}{(n^2 - 1) + \frac{\beta^2}{\alpha^2}}. \quad (24)$$

It is evident from (24) that for all values of n and β/α which are larger than one, $\rho_2 < 1$. In order also to satisfy the condition $0 < \rho_2$, it is necessary to require that

$$n^2 > \frac{\beta^2}{\alpha^2} + 1.$$

Furthermore, all the angles defined by (23) must be real if the Hankel function approximations are to be valid, a requirement that reduces to the single additional condition,

$$n^2 < \left(\frac{\beta}{\alpha} + 1 \right)^2.$$

The significant physical constants of the dielectric clad cylinder must therefore have values that are in accord with

$$\frac{\beta^2}{\alpha^2} + 1 < n^2 < \left(\frac{\beta}{\alpha} + 1 \right)^2. \quad (25)$$

The clear implication of this inequality is that only for certain combinations of dielectric thickness and refractive index will the ray represented by (21) be scattered back in the direction of incidence.

In the region defined by (25), the solutions of (22) are real and may be used for the asymptotic evaluation of (21). The result of this calculation is

$$\begin{aligned} \psi_2^{\text{sc.}} = & \frac{1}{(n^2 - 1)^2} \left[(n^2 - 1)^2 - \frac{\beta^4}{\alpha^4} \right] \left(\frac{B}{2k_0 r} \right)^{1/2} \\ & \cdot \exp i \left[k_0 r - \alpha_0 \left(n^2 - 3 \frac{\beta^2}{\alpha^2} + 1 \right) \right], \end{aligned} \quad (26)$$

where

$$\frac{1}{B} = \frac{1}{\alpha(\sin \zeta)_2} - \frac{1}{\beta(\sin \eta)_2} + \frac{1}{\beta_0(\sin \eta_0)_2}.$$

The physical optics interpretation of (26) is not obvious but the contribution clearly includes more than the straight, axial ray which is reflected at the conducting cylinder. It seems likely that other internally reflected rays are also involved and that these would be consistent with the restrictions of (25). Such a ray model needs to be elucidated by further calculations.

Modes in Rectangular Guides Partially Filled with Transversely Magnetized Ferrite*

G. BARZILAI† AND G. GEROSA†

Summary—A general modal solution for a rectangular guide partially filled with a slab of ferrite transversely magnetized and situated against one side wall is considered. The relative characteristic equation, which has been obtained in a previous work, is numerically solved for some typical cases and for modes of zero, first and second order with respect to the dependence along the direction of the dc magnetic field. The results are summarized in fifteen diagrams giving the propagation constant vs the ferrite slab thickness.

From the preceding analysis it can be concluded that:

- 1) propagating modes of any order may exist;
- 2) modes of zero order are not always the lowest propagating modes, in the sense that higher order modes can propagate when zero order modes are attenuated;
- 3) cases can exist when all the unattenuated propagating modes travel in the same sense.

INTRODUCTION

IT IS the purpose of this work to discuss in some detail the modal spectrum for a rectangular guide partially filled with a slab of ferrite transversely magnetized and situated against one side wall.

The characteristic equation for such a structure was derived by Kales, Chait, and Sakiotis,¹ Lax, Button, and Roth,^{2,3} and others, by assuming no dependence along the direction of the dc magnetic field. In a previous work,⁴ we have derived the characteristic equation for the general case, *i.e.*, with no restriction on the dependence along the direction of the dc magnetic field.

To our knowledge, all the discussions of the characteristic equation available up to date are those relative to modes with no dependence along the direction of the dc magnetic field. It is, however, interesting to discuss the general case, particularly in view of the fact that it can happen that in certain conditions modes with no dependence along the direction of the dc magnetic field cannot propagate, while other modes can.

We shall assume that the reader is acquainted with our previous work⁴ and we shall only recall here the symbols used. The structure considered is the one indicated in Fig. 1.

* The investigation reported herein has been made possible by the support of the U. S. Air Force, under Contract No. AF-61(052)-101.

† Electrotechnical Institute, University of Rome, Rome, Italy.

¹ M. L. Kales, H. N. Chait, and N. G. Sakiotis, "A nonreciprocal microwave component," *J. Appl. Phys.*, vol. 24, pp. 816-817; June, 1953.

² B. Lax, K. J. Button, and L. M. Roth, "Ferrite phase shifters in rectangular waveguide," *J. Appl. Phys.*, vol. 25, pp. 1413-1421; November, 1954.

³ K. J. Button and B. Lax, "Theories of ferrites in rectangular waveguides," *IRE TRANS. ON ANTENNAS AND PROPAGATION*, vol. AP-4, pp. 531-537; July, 1956.

⁴ G. Barzilai and G. Gerosa, "Modes in rectangular guides filled with magnetized ferrite," *Il Nuovo Cimento*, vol. X-7, pp. 685-697; March, 1958.

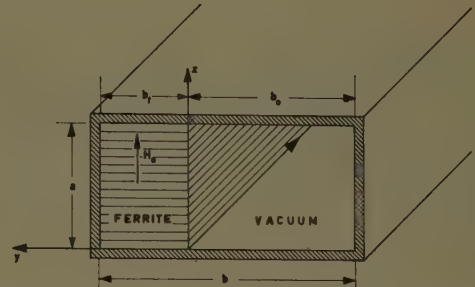


Fig. 1—Geometry of the rectangular guide partially filled with ferrite.

We shall assume for the ferrite region the following tensor permeability:

$$\underline{\mu} = \mu_0 \begin{vmatrix} \mu_1 & j\mu_2 & 0 \\ -j\mu_2 & \mu_1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

where

$$\mu_1 = 1 + \frac{\rho}{1 - \tau^2}; \quad \mu_2 = \frac{\tau\rho}{1 - \tau^2}; \quad \rho = \frac{M_0}{\mu_0 H_0};$$

$$\tau = \frac{\omega}{\omega_0}$$

M_0 is the saturation magnetization, ω is the circular applied frequency (time dependence $\exp[j\omega t]$ is assumed), $\omega_0 = -\gamma H_0$ is the circular resonant frequency, γ is the gyromagnetic ratio for the electron, H_0 is the internal dc magnetic field, and μ_0 is the permeability of the vacuum. The dielectric constant of the ferrite region is assumed to be $\epsilon_0\epsilon$, where ϵ_0 is the dielectric constant of the vacuum. Lengths will be measured by assuming as unit $1/\omega\sqrt{\mu_0\epsilon_0}$. Spatial dependence of the form $\exp[j(k_x x + k_y y + k_z z)]$ will be assumed.

For a prescribed pair k_x, k_y in the ferrite region, there exist two values of k_y^2 , namely k_{y1}^2 and k_{y2}^2 , while in the vacuum region there exists only one value of k_y^2 , namely k_{y0}^2 .

By writing a general modal solution for the structure indicated in Fig. 1, we have obtained⁴ a transcendental characteristic equation of the form:

$$f(k_x, k_z) = 0 \quad (1)$$

where

$$k_z = \frac{m\pi}{a} \quad (m = 0, 1, 2, \dots)$$

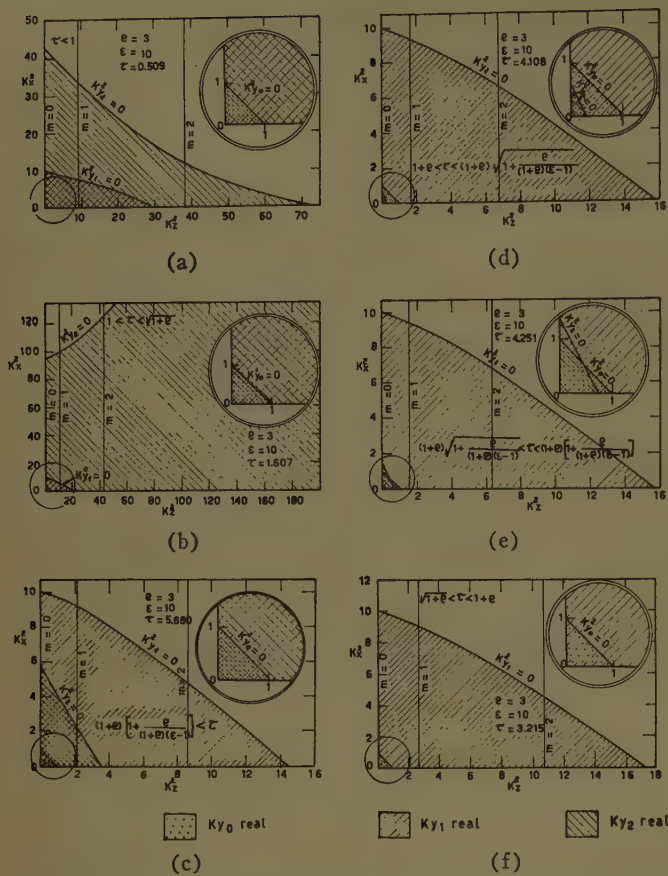


Fig. 2—Zones of the first quadrant of the k_x^2, k_z^2 plane in which k_{y1} , k_{y2} and k_{y0} are real or imaginary. (a) The curve $k_{y2}^2=0$ lies above $k_{y1}^2=0$ and the corresponding asymptote has a negative slope; (b) the curve $k_{y2}^2=0$ lies above $k_{y1}^2=0$ and $k_{y0}^2=0$ and the corresponding asymptote has a positive slope; (c) the curve $k_{y2}^2=0$ does not cross the first quadrant of the k_x^2, k_z^2 plane; (d) the curve $k_{y2}^2=0$ lies below $k_{y1}^2=0$ and $k_{y0}^2=0$; (e) the curve $k_{y2}^2=0$ lies below $k_{y1}^2=0$ and crosses $k_{y0}^2=0$; (f) the curve $k_{y2}^2=0$ lies between $k_{y1}^2=0$ and $k_{y0}^2=0$.

In what follows, we shall call those modes corresponding to $m=0$, modes of zero order, those corresponding to $m=1$, modes of first order, and so on. Further details on the present work may be obtained in another report.⁵

SOLUTION REGIONS

There are several parameters which determine the solutions of (1), namely: quantities characteristic of the ferrite medium, *i.e.*, M_0 and ϵ ; quantities describing the structure, *i.e.*, a , b and b_f ; and impressed quantities, *i.e.*, ω and H_0 . The three quantities M_0 , ω and H_0 enter our problem through the adimensional parameters ρ and τ , so that the actual parameters to be considered are ρ , ϵ , τ , a , b and b_f .

For a given mode order, (1) determines the relative propagation constants k_x . In what follows, we shall only look for real values of k_x , *i.e.*, for unattenuated propagating modes.

⁵ G. Barzilai and G. Gerosa, "Modes in Rectangular Guides Partially Filled with Transversely Magnetic Ferrite," Electrotech. Inst., Inst. of Rome, Rome, Italy, Tech. Note No. 1, Contract No. AF 61(052)-101; June 3, 1959.

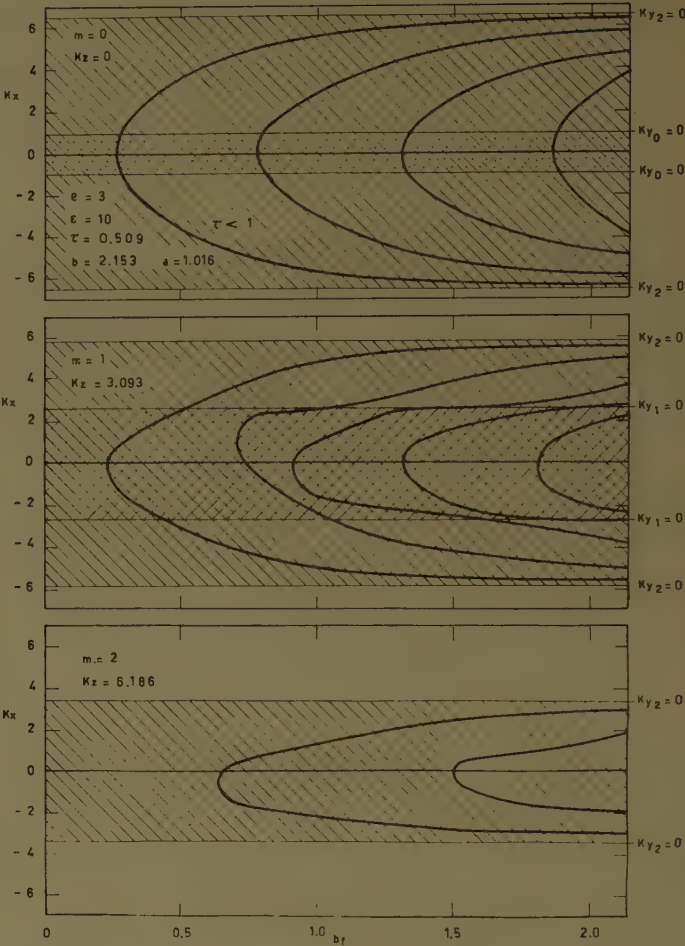


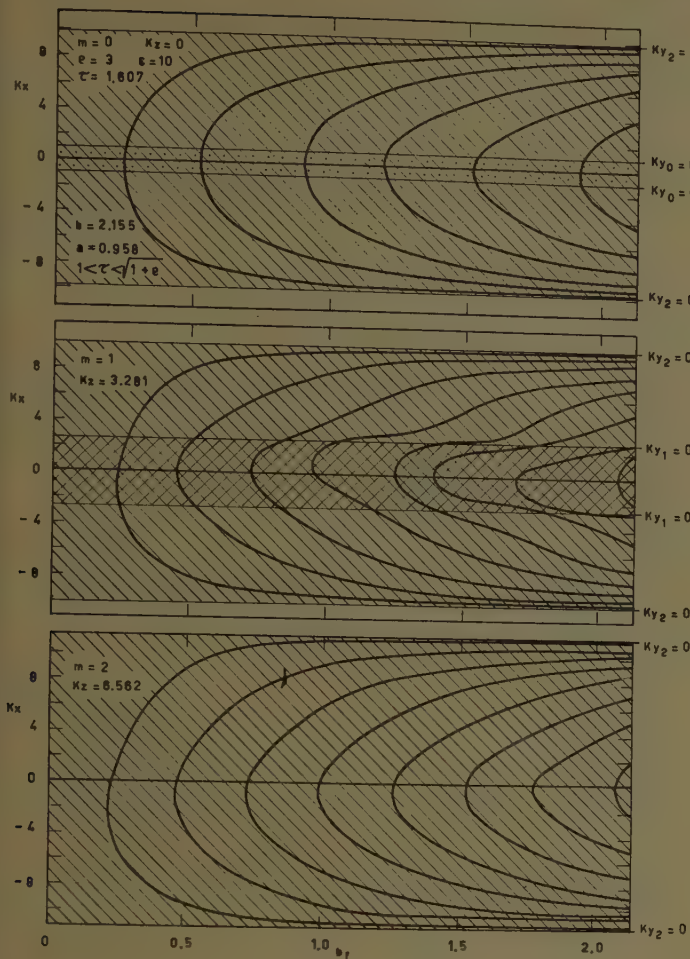
Fig. 3—The mapping of the solutions of the characteristic equation for the range of τ indicated. The curves have been calculated for the indicated values of ρ , ϵ , τ , b , and a . These values may be taken to correspond to:

$$M_0 = 0.3 \frac{Wb}{m^2}; \quad H_0 = \frac{10^6}{4\pi} \text{ A/m}; \quad f = \frac{\omega}{2\pi} = 1425 \text{ mc};$$
$$\frac{b}{\omega \sqrt{\mu_0 \epsilon_0}} = 2.84 \text{ inch}; \quad \frac{a}{\omega \sqrt{\mu_0 \epsilon_0}} = 1.34 \text{ inch}.$$

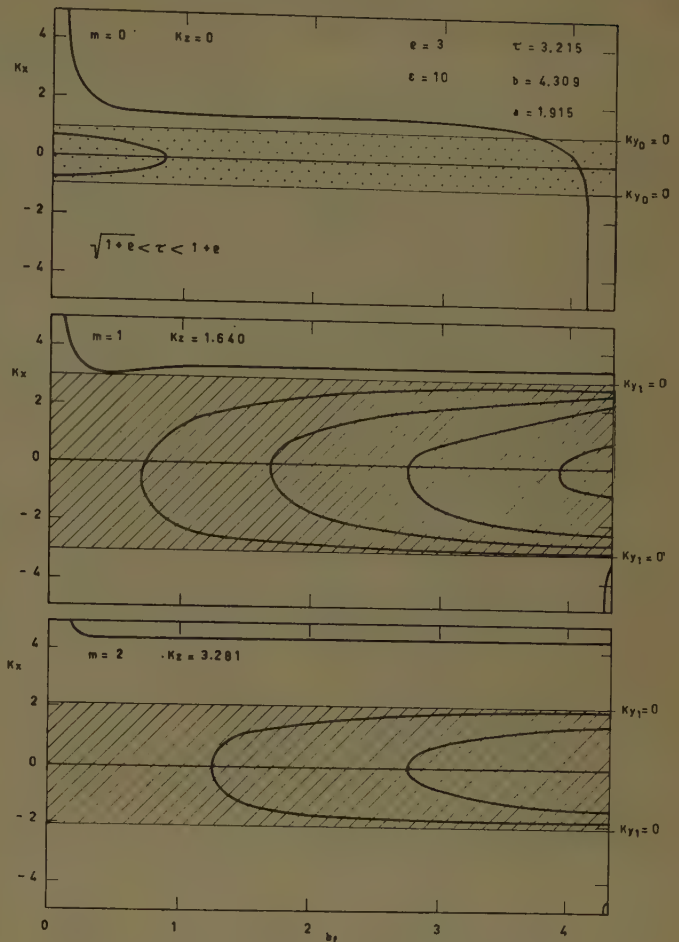
With respect to the field configuration in the cross section of the guide, the mode order determines the z dependence. The y dependence is determined by the values of k_{y1} , k_{y2} , and k_{y0} , which can assume real or purely imaginary values, being k_x and k_z real.

With reference to the first quadrant of the k_x^2, k_z^2 plane, the regions where k_{y1} , k_{y2} , and k_{y0} are real or imaginary are delimited by the curves $k_{y1}^2=0$; $k_{y2}^2=0$, which have been shown to belong to an hyperbola, and $k_{y0}^2=0$, which is easily realized to be a 45° straight line crossing the k_x^2 and k_z^2 axes at the points (0.1) and (1.0) by virtue of the normalization assumed.

By keeping ρ fixed and letting τ vary, there are six typical cases according to the mutual configurations of the three mentioned curves. We start by noting that the straight line $k_{y0}^2=0$ does not vary with τ . The curve $k_{y1}^2=0$ intercepts the k_x^2 axis at the point $k_x^2=\epsilon$; it has an asymptote having a slope -1 and intercepting the k_z^2 axis at the point $k_z^2=\epsilon(1+\rho)$; we can conclude therefore that this curve does not change much as τ

Fig. 4—The same as Fig. 3 except for $f=4500$ mc;

$$\frac{b}{\omega\sqrt{\mu_0\epsilon_0}} = 0.9 \text{ inch}; \quad \frac{a}{\omega\sqrt{\mu_0\epsilon_0}} = 0.4 \text{ inch}.$$

Fig. 5—The same as Fig. 3 except for $f=9000$ mc;

$$\frac{b}{\omega\sqrt{\mu_0\epsilon_0}} = 0.9 \text{ inch}; \quad \frac{a}{\omega\sqrt{\mu_0\epsilon_0}} = 0.4 \text{ inch}.$$

varies. The curve $k_{y2}^2=0$ intercepts the k_x^2 axis at the point $k_x^2 = \epsilon(\mu_1^2 - \mu_2^2)/\mu_1$ and has an asymptote which has a slope $-1/\mu_1$ and intercepting the k_x^2 axis at the point $k_x^2 = \epsilon[1 + \mu_2^2/(\mu_1^2 - \mu_1)]$. This curve therefore changes considerably as τ varies.

With reference to the position of the curve $k_{y2}^2=0$ with respect to $k_{y1}^2=0$ and $k_{y0}^2=0$, we shall divide the field of variability of τ from 0 to ∞ into six regions, which are indicated in Fig. 2. The relative diagrams have been drawn assuming suitable numerical values for the parameters.

In the six diagrams of Fig. 2, the various zones delimited by the curves $k_{y1}^2=0$, $k_{y2}^2=0$ and $k_{y0}^2=0$ have been shaded in different ways in order to recognize the zones where k_{y1} , k_{y2} , and k_{y0} are real or imaginary. It is understood that when one of the three typical shadings indicated is present, the relative k_y is real and when it is not, it is imaginary. For instance, when no shading exists, the three k_y are all imaginary; when all three shadings are present the three k_y are all real, and so on.

NUMERICAL ANALYSIS

For each of the six regions in which the field of variability of τ has been divided, we have chosen the nu-

merical value of τ indicated in Fig. 2 and convenient values of a and b .

By numerically solving (1), we have drawn the diagrams of k_x vs b_1 for modes of zero, first and second order. The relative values of k_x are indicated by vertical straight lines drawn in the various cases of Fig. 2. In Figs. 3–7, we have drawn the solution curves for the cases of Fig. 2(a)–2(c) and 2(e). The solution curves for the remaining two cases are similar to those of case (e) and have been omitted for brevity.

In Figs. 3–7, the shadings correspond to those discussed with reference to Fig. 2. It should be noted, however, that for $m=0$, (1) breaks down into two equations corresponding to TE and TM modes associated with k_{y2} and k_{y1} , respectively; but since TM modes cannot exist in the rectangular guide, the shading relative to k_{y1} has been omitted in the diagrams relative to zero order modes.

The numerical solutions have been obtained by using an electronic digital computer. Asymptotic analyses were carried out, when necessary, in order to complete the diagrams.

Inspection of the diagrams found shows that, except

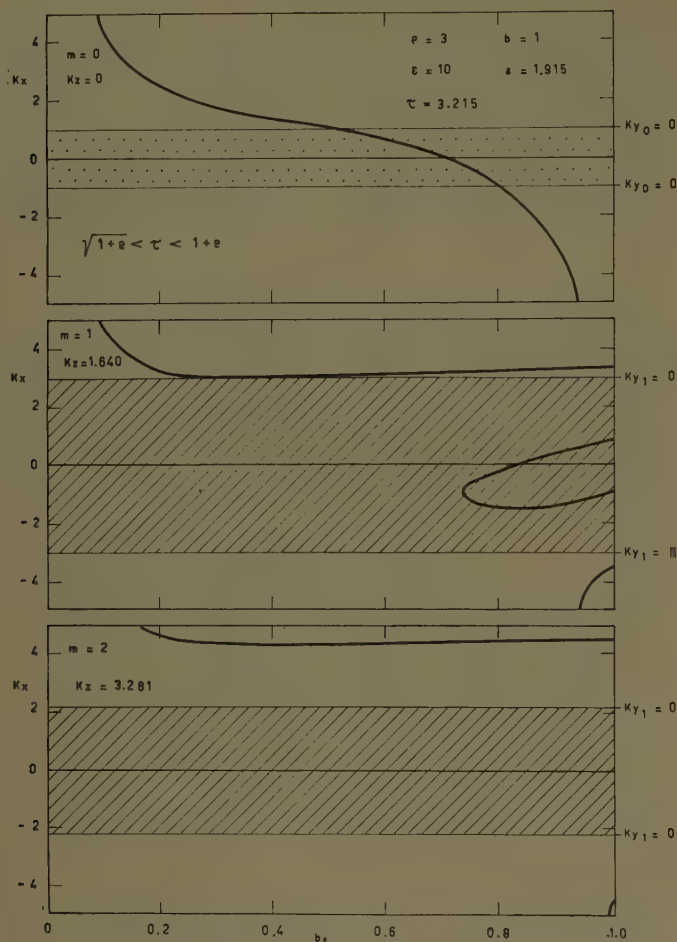


Fig. 6—The same as Fig. 5 except for

$$\frac{b}{\omega\sqrt{\mu_0\epsilon_0}} = 0.209 \text{ inch.}$$

for $b_f=0$ and $b_f=b$ when the characteristic equation contains only even powers of k_x , the solutions are not reciprocal, in the sense that a solution $+k_x$ does not necessarily imply the solution $-k_x$.

The numerical analysis carried out, while necessarily incomplete, suggests the following considerations on the general behavior of the modal spectrum, which in the light of the results obtained seem reasonable.

For the cases of Figs. 3 and 7, the solution curves of nonzero order modes decrease in number as m increases. Therefore, above a certain order mode, propagation cannot exist. With reference to our previous work,⁴ this has already been shown for the completely filled guide.

In contrast with this behavior, the cases of Figs. 4–6 show that propagating modes of any order can exist. For Fig. 4, this is true for b_f if it is not too small, while for Figs. 5 and 6, this is true for any value of b_f by virtue of the asymptotic behavior of some of the solution curves, which have as asymptotes the straight vertical lines $b_f=0$ and $b_f=b$.

In the case of Fig. 4, by sufficiently reducing b , modes of zero order go below cutoff, while higher order modes still propagate. This has been shown by us for $b=b_f$ and it seems to be so even for $b_f < b$.

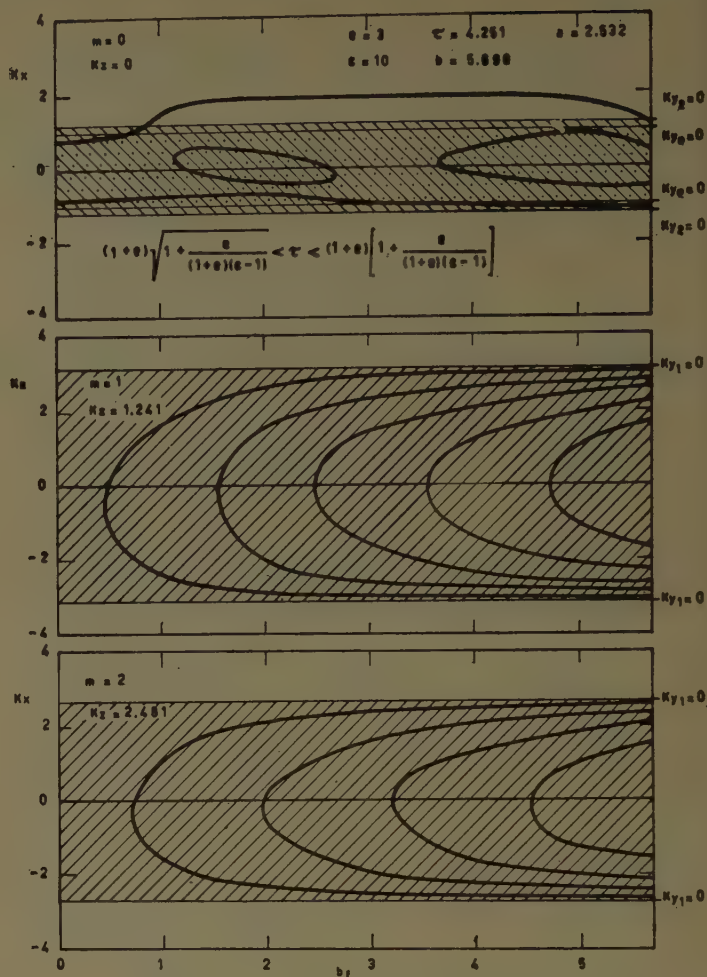


Fig. 7—The same as Fig. 3 except for

$$= 11,900 \text{ mc; } \frac{b}{\omega\sqrt{\mu_0\epsilon_0}} = 0.9 \text{ inch; } \frac{a}{\omega\sqrt{\mu_0\epsilon_0}} = 0.4 \text{ inch.}$$

The diagrams of Fig. 5 show that when modes of zero order cannot propagate in both directions, higher order modes can.

With reference to Figs. 5 and 6, it is apparent that by narrowing the width b of the guide and leaving the other parameters unchanged, for suitable values of b_f the structure becomes unidirectional for modes of any order.

CONCLUSIONS

The modal spectrum for a rectangular guide partially filled with a slab of ferrite situated against one side wall and transversely magnetized has been discussed.

The principal conclusions are the following:

1) There are regions of variability of τ , i.e., of the ratio between the applied and resonant frequencies, for which modes of any order can propagate.

2) Zero order modes are not always the lowest modes, in the sense that higher order modes can propagate when zero order modes cannot.

3) It is possible to find unidirectional structures, i.e., structures in which, under the hypotheses by which the characteristic equation is assumed to be derived, all the unattenuated propagating modes have the phase velocity in the same sense.

Asymptotically Expansible Solutions of the Helmholtz Equation

KRYSTYN BOCHENEK†

ABSTRACT

Much attention has been paid in recent years to expansions in the form,

$$u = e^{ikL} \sum_{\nu}' \frac{A_{\nu}}{(ik)^{\nu}}. \quad (1)$$

By means of these expansions many problems of diffraction have been investigated.¹

Though in many particular cases, it has been found that the series of this form are the asymptotic ones, our present knowledge concerning the whole family of solutions admitting asymptotic expansions is still unsatisfactory. From among recent works on this subject the author will mention a very interesting one² concerning expansions of a similar type.

In this report the writer proposes to present the results of his as yet unpublished investigations of a rather mathematical character. In this summary only outlines of the results are given.

† Polish National Committee of URSI, Warsaw, Poland.

¹ J. B. Keller, R. M. Lewis, and B. D. Seckler, *Commun. Pure Appl. Math.*, no. 2; 1956.

² R. M. Lewis, New York University, New York, N. Y., Res. Rept. MME-8; 1957.

Let us consider a function,

$$u = k^{1/2} \int_L a(\alpha) e^{ik[r \cos(\theta - \alpha) + b(\alpha)]}, \quad (2)$$

where $a(\alpha)$ and $b(\alpha)$ are analytic functions; L is a contour in the complex plane α ; r and θ , polar coordinates.

This expression represents a superposition of plane waves. Given an infinite domain D situated outside of a given curve C in a plane r, θ , it is possible under appropriate conditions to prove that the function u satisfies the Helmholtz equation in the whole domain D and the radiation condition at the infinity, and has an asymptotic expansion of the type (1). This expansion can be differentiated term by term. Because of this we are sure that the coefficients A_{ν} satisfy the recursive system of equations describing the propagation of those coefficients along the rays of geometrical optics.

It is interesting to note that in the case when an asymptotic expansion of (2) is computed by means of the saddle-point method, the eiconal will be constructed in a classical manner as an envelope of a family of solutions formed by means of a complete integral. The coefficients of expansion A_{ν} are effectively obtained.

INDEX TO SYMPOSIUM ON ELECTROMAGNETIC THEORY

SPECIAL SUPPLEMENT TO IRE TRANSACTIONS ON
ANTENNAS AND PROPAGATION

Volume AP-7, December, 1959

Index to Authors

A

Adachi, S. Dec S406
Allen, C. C. Dec S387

B

Bachynski, M. P. Dec S226, S337
Bailin, L. L. Dec S458
Barlow, H. M. Dec S147
Barzilai, G. Dec S471
Bazer, J. Dec S12
Bochenek, K. Dec S475
Bracey, M. F. Dec S219
Braunbek, W. Dec S71
Bremmer, H. Dec S175
Bresler, A. D. Dec S261
Brown, A. Dec S12
Brown, J. Dec S169
Brysk, H. Dec S330

C

Chisholm, R. N. Dec S279
Chu, T. S. Dec S273
Clemmow, P. C. Dec S7
Cullen, A. L. Dec S219

D

Deschamps, G. A. Dec S371

F

Felsen, L. B. Dec S231
Franz, W. Dec S68
Friedman, B. Dec S227
Furutsu, K. Dec S209

G

Gardner, C. S. Dec S87
Gerosa, G. Dec S471
Gillespie, E. F. F. Dec S219
Goodrich, R. F. Dec S28
Goubau, G. S140

H

Hansen, R. C. Dec S458
Hedgecock, N. E. Dec S284
Held, G. Dec S296
Hessel, A. Dec S201
Hewish, A. Dec S120
Hoffman, W. C. Dec S301

J

Johnston, T. W. Dec S337

K

Karal, F. C., Jr. Dec S91
Karbowski, A. E. Dec S191
Karp, S. N. Dec S91
Kay, I. Dec S255
Kazarinoff, N. D. Dec S21
Keller, J. B. Dec S87
Keys, J. E. Dec S77
King, R. Dec S440
Klante, K. Dec S68
Knudsen, H. L. Dec S361
Kodis, R. D. Dec S468
Kouyoumjian, R. G. Dec S273, S379, S406

L

Levis, C. A. Dec S424
Levy, B. R. Dec S52

M

Maclean, T. S. M. Dec S379
Marie, P. Dec S412
McCormick, G. C. Dec S288
McLay, A. B. Dec S284
Meixner, J. Dec S435
Mittra, R. Dec S244

N

Neugebauer, H. E. J. Dec S226
Noble, B. Dec S37
Northover, F. H. Dec S340

O

Oliner, A. A. Dec S201
Olte, A. Dec S61

P

Papadopoulos, V. M. Dec S78
Peake, W. H. Dec S324
Piefke, G. Dec S183
Polk, C. Dec S414
Primich, R. I. Dec S77

R

Ritt, R. K. Dec S21
Robieux, J. Dec S118

Ronchi, L. Dec S125
Rumsey, V. H. Dec S103, S117
Russo, V. Dec S125
Ryle, M. Dec S120

S

Saxon, D. S. Dec S320
Schelkunoff, S. A. Dec S133
Seracchioli, F. Dec S424
Shakeshaft, J. R. Dec S120
Shkarofsky, I. P. Dec S337
Shmoys, J. Dec S88
Silver, S. Dec S61
Sinclair, G. Dec S402
Staniforth, J. A. Dec S219

T

Toraldi di Francia, G. Dec S125
Twersky, V. Dec S307
Tyras, G. Dec S296

V

Van Bladel, J. Dec S119
Van Sickle, R. G. Dec S406

W

Wait, J. R. Dec S132, S154
Weston, V. H. Dec S43

Index to Subjects

A

Air, High-Temperature, Electromagnetic Properties of: Dec S337
Anisotropic Layers, Propagation of Electromagnetic Waves Through: Dec S296
Antennas:
 Broad-Band Multislot: Dec S412
 on Circular Cylinders: Dec S361
 Equiangular Spiral: Dec S117
 Finite Conical: Dec S406
 Helical, Bandwidth of: Dec S379
 Loaded Biconical, Resonance and Super-gain Effects in: Dec S414
 Pattern Calculations, Numerical Integration Methods for: Dec S387
 and Scattering Problems, Numerical Solution of: Dec S402
Aperture, Circular, Diffraction of Scalar Waves by: Dec S12
Arrays, Linear: Currents, Impedances, and Fields of: Dec S440
Arrays, Long End-Fire Dipole, Calculated Phase Velocity of: Dec S424
Attenuation in Wedge and Septate Waveguides: Dec S279

B

Bandwidth of Helical Antennas: Dec S379
Boundary Value Problem and Finite Range Wiener-Hopf Integral Equation in a Waveguide: Dec S244

C

Caustic, Fields in the Neighborhood of: Dec S255
Cavity, Ferrite-Filled Cylindrical: Dec S273
Cone and Wedge Surfaces with Varying Surface Impedance, Properties of: Dec S231

D

Decay Exponents and Diffraction Coefficients for Surface Waves: Dec S52
Dielectric Plates, Parallel Matching of, to Free Space: Dec S288
Diffraction:
 Coefficients and Decay Exponents for Surface Waves: Dec S52
 by Disks and Strips, Equations for High-Frequency: Dec S37
 by a Half-Plane with a Special Impedance Variation: Dec S88
 of Nearly Plane Waves by 45° and 90° Conduction Wedges: Dec S284
 of a Plane Wave by a Funnel-Shaped Screen: Dec S71
 Scalar, by an Elliptic Cylinder: Dec S21
 of Scalar Waves by a Circular Aperture: Dec S12
 by Smooth Conical Obstacles: Dec S226
 by Surfaces of Variable Curvature: Dec S68

Theory, Infinite Integral Transforms in: Dec S7

Dipole Arrays, Long End-Fire, Calculated Phase Velocity of: Dec S424
Dipole, Pulsed, in an Interface, Field of: Dec S87
Discontinuity Problem at the Input to an Anisotropic Waveguide: Dec S261

E

Electromagnetic Field in a Randomly Inhomogeneous Medium: Dec S301
Electromagnetic Waves with Some Natural Surfaces, Interaction of: Dec S324
Excitation of the Waves of Proper Solutions: Dec S209

F

Far-Field Scattering from Simple Shapes, Experimental Determination of: Dec S77
Far Fields of a Line Source at the Tip of an Absorbing Wedge: Dec S91
Ferrite-Filled Cylindrical Cavity: Dec S273
Field Analysis, Near, New Method of: Dec S458
Field, Electromagnetic, in a Randomly Inhomogeneous Medium: Dec S301
Fields in the Neighborhood of a Caustic: Dec S255
Fock Theory: Dec S28

G

- Gases, Ionized, Propagation of Electromagnetic Waves in: Dec S340
 Guide, Transmission Characteristics of: Dec S183
 Guided Waves and Radiation: Dec S191
 Guided Waves on Sinusoidally-Modulated Reactance Surfaces: Dec S201
 Guiding of Waves by Uniformly Rough Surfaces: Dec S154

H

- Helical Antennas, Bandwidth of: Dec S379
 Helmholtz Equation, Asymptotically Expandable Solutions of: Dec S475
 Helmholtz's Theorem in Finite Regions: Dec S119
 Huygens' Principle, New Forms of: Dec S103

I

- Impedance Properties of Complementary Multiterminal Planar Structures: Dec S371
 Interfaces, Waves on: Dec S140
 Interferometers, Radio, Synthesis of Large Radio Telescopes by Use of: Dec S120
 Ionized Gases, Propagation of Electromagnetic Waves in: Dec S340

L

- Linear Systems, Network Theory and Its Relation to the Theory of: Dec S435

M

- Matching of Parallel Dielectric Plates to Free Space: Dec S288
 Meteor Trails, High-Density, Scattering by: Dec S330
 Mirror, Stepped Zone, for Microwaves; Experimental Test of: Dec S125
 Modes in Rectangular Waveguides Partially Filled with Transversely Magnetized Ferrite: Dec S471

N

- Near Field Analysis, Near Method of: Dec S458

- Network Theory and Its Relation to the Theory of Linear Systems: Dec S435

P

- Phase Velocity, Calculated, of Long End-Fire Dipole Arrays: Dec S424
 Planar Structures, Complementary Multiterminal, Impedance Properties of: Dec S371
 Propagation of Electromagnetic Waves Through Anisotropic Layers: Dec S296
 Propagation of Electromagnetic Waves in Ionized Gases: Dec S340
 Propagation and Scattering of Electromagnetic Waves, Modified WKB Methods for: Dec S320
 Pulse Return from a Sphere: Dec S43
 Pulses, Diffraction and Refraction of: Dec S78

R

- Radiation and Guided Waves: Dec S191
 Radio Telescopes, Synthesis of, by Use of Radio Interferometers: Dec S120
 Reactance Surfaces, Sinusoidally-Modulated, Guided Waves on: Dec S201
 Refraction and Diffraction of Pulses: Dec S78
 Resonance and Supergain Effects in Loaded Biconical Antennas: Dec S414

S

- Scattering of Waves:
 and Antenna Problems, Numerical Solution of: Dec S402
 BackScattering from Cones and Spheroids: Dec S67
 Back Scattering at High Frequencies from a Conducting Cylinder with Dielectric Sleeve: Dec S468
 Far-Field, Experimental Determination of: Dec S77
 by High-Density Meteor Trails: Dec S330
 and Propagation of Electromagnetic Waves, Modified WKB Methods for: Dec S320
 by Quasi-Periodic and Quasi-Random Distributions: Dec S307

- Stepped Zone Mirror for Microwaves, Experimental Test of: Dec S125
 Spiral Equiangular Antenna: Dec S117
 Supergain and Resonance Effects in Loaded Biconical Antennas: Dec S414
 Surface Waves:

- Anatomy of: Dec S133
 Decay Exponents and Diffraction Coefficients for: Dec S52
 Launchers, Theoretical Results for: Dec S169
 Over a Lossy Conductor: Dec S227
 Papers, Preface to: Dec S132
 Propagation Trajectories Associated with the Sommerfeld Problem: Dec S175
 Research in Sheffield: Dec S219
 Supported by Cylindrical Surfaces: Dec S147

T

- Transmission Characteristics of a Guide: Dec S183
 Transmission Coefficient from a Transmitting to a Receiving System, Theorems on: Dec S118

W

- Waveguides:
 Anisotropic, Discontinuity Problem at the Input to: Dec S261
 Finite Range Wiener-Hopf Integral Equation and a Boundary Value Problem in: Dec S244
 Modes in, Partially Filled with Ferrite: Dec S471
 Wedge and Septate, Attenuation in: Dec S279
 Waves on Interfaces: Dec S140
 Wedges and Cone Surfaces with Varying Surfaces Impedance, Properties of: Dec S231
 Wiener-Hopf Integral Equation, Finite Range, and a Boundary Value Problem in a Waveguide: Dec S244
 WKB Methods, Modified, for the Propagation and Scattering of Electromagnetic Waves: Dec S320

AMCI**TYPE 1026**

SLOTTED LINES

for
accurate
measurement of
impedance in rigid
and flexible coaxial
transmission lines.

FEATURES:

- Rated residual VSWR under 1.01
- Rated error in detected signal under 1.005
- Available in 20, 40, 60, 80, and 130 inch lengths.

Write for
Bulletin E-958D

The outer conductor of the Type 1026 Slotted Lines is made of two substantial aluminum castings, carefully machined and dowelled together, with the important surfaces finished by a hand scraping operation. The inner conductor is ground to a close tolerance, supported by compensated dielectric pins, and longitudinally positioned by a compensated dielectric anchor at the feed end.

AMCI Tapered Reducers, Instrument Loads, and Impedance Standard Lines are available for use with the Type 1026 Slotted Lines in making measurements of a wide range of rigid and flexible coaxial lines.



ANTENNA SYSTEMS—COMPONENTS—AIR NAVIGATION AIDS—INSTRUMENTS

ALFORD

Manufacturing Co.
299 ATLANTIC AVE. BOSTON, MASS.



In antenna systems
KENNEDY capability is

total capability

TOTAL capability in the field
of antenna systems?

It's the capability to do the basic R & D in microwave propagation . . . to design and develop the antenna system . . . to manufacture the dish, the mount, and all waveguide components, horns, etc. . . . to provide complete field engineering service which includes site surveying, construction and erection, final checkout, and servicing.

In short, it's the capability to do it all — a total service from a single source.

Opportunities for outstanding scientists and engineers in this fast-growing organization.

ANTENNA EQUIPMENT**D. S. KENNEDY & CO.**

COHASSET, MASS., Evergreen 3-1200

West Coast Affiliate . . .

SATELLITE KENNEDY INC. of CALIFORNIA
P.O. Box 1711, Monterey, Cal., FR 3-2461

Down-to-earth SOLUTIONS to out-of-this-world PROBLEMS . . . Tracking Antennas—Radio Telescopes—Radar Antennas—"Trans-Horizon" Antennas—Tropospheric Scatter—Ionospheric Scatter

For Information

Concerning

Advertising

Rates

Contact

Mr. Delmer C. Ports

Jansky and Bailey, Inc.

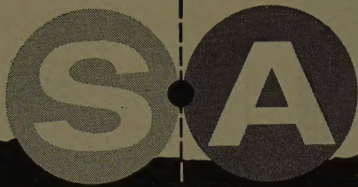
1339 Wisconsin Ave.,

N.W.

Washington 7, D. C.

Telephone:

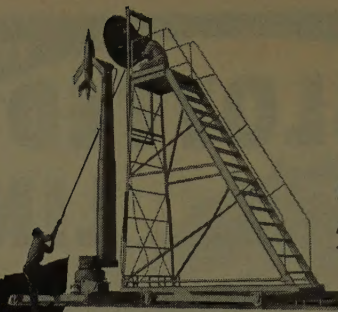
Federal 3-4800



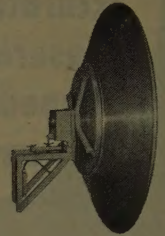
Complete

**ANTENNA
PATTERN**

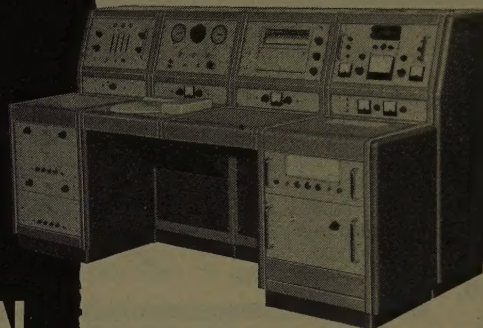
INSTRUMENTATION



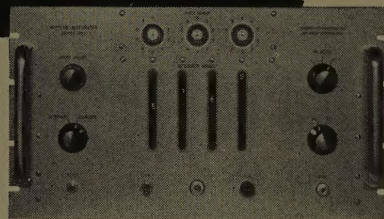
MODEL RANGE
AND TRANSMITTING
TOWERS



STANDARD
GAIN HORNS
AND TRANSMITTING
ANTENNAS



CENTRALIZED
OPERATING
CONSOLES



FOURIER
INTEGRAL
COMPUTERS



ANTENNA PATTERN
INTEGRATORS

S-A Specializes

Leaders in antenna test instrumentation, S-A furnishes the antenna design engineer with every research tool he needs. Instruments include polar and rectangular recorders, wide range receivers, antenna positioners, model range towers, remote tuned signal sources, transmitting and standard gain antennas, microwave components, and Fourier integral computers.

If you would like complete details, write for our new data file or call your nearby S-A engineering representative whose name and phone number are listed below. He's a good man to know.



**SCIENTIFIC-
ATLANTA, INC.**

2162 PIEDMONT ROAD, N. E. • ATLANTA 9, GEORGIA
TRinity 5-7291



ENGINEERING REPRESENTATIVES

Atlanta 9, Georgia: Scientific Sales Engineering Company, TRinity 3-2475 • Baltimore (Towson), Maryland: Gawler-Knoop Company, VALley 5-3151 • Boston (Waltham), Massachusetts: George Gostenhofer & Associates, Inc., TWInbrook 4-9500 • Chicago 45, Illinois: Pivan Engineering Company, BRlrgate 4-9135 • Cleveland, Ohio: The Tiby Company, ERievew 1-5335 • Dayton 6, Ohio: The Tiby Company, CRestview 7-3822 • Denver (Loveland), Colorado: Kelley Enterprises, NORmandy 7-1376 • Fort Worth, Texas: Carey-Wolf Company, PERshing 8-1702 • Huntsville, Alabama: Scientific Sales Engineering Company, JEFFerson 9-5552 • Indianapolis, Indiana: Pivan Engineering Company, CLifford 3-0444 • Kansas City 14, Missouri: LeeMark Associates, NighTingale 2-3313 • Los Angeles (San Gabriel), California: J. T. Hill Company, ATLantic 7-9633 • New York City (Roseland, New Jersey): Gawler-Knoop Company, CAldwell 6-4545 • Orlando, Florida: Scientific Sales Engineering Company, GA 4-0730 • Philadelphia (Wyncote), Pennsylvania: Gawler-Knoop Company, WAVERly 7-1820 • Phoenix, Arizona: J. T. Hill Company, CRestwood 7-0506 • San Diego 7, California: J. T. Hill Company, ACAdemy 3-7133 • San Francisco (San Carlos), California: J. T. Hill Company, LYtell 3-7693 • Seattle, Washington: Rush S. Drake Associates, EAST 3-8545 • St. Louis, Missouri: LeeMark Associates, MISSION 7-1470 • St. Petersburg, Florida: Scientific Sales Engineering Company, PHOne: 5-7874 • Syracuse Medical Equipment Div., HUDson 5-8621 • Washington, D. C. (Silver Spring, Md.): Gawler-Knoop Company, JUNiper 5-7550 • Winston-Salem, North Carolina: Scientific Sales Engineering Company, PARk 3-3281 • EXPORT (New York City): Szucs International Corporation, WHItehall 4-7959.

**Because of continuous expansion
DORNE AND MARGOLIN, INC.**

always welcomes inquiries
from engineers and
scientists
interested
in:

Antennas
Microwave Circuits
Countermeasures

Our program, which ranges from
research to production, will
interest the scientifically
oriented individual seeking
congenial and rewarding
industrial employment.

We offer all the usual benefits
including full tuition
reimbursement, plus
in-the-house classes
and seminars.

Stock participation, although not
offered to the public,
is available to all
employees.

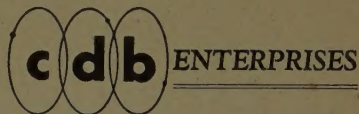
DORNE AND MARGOLIN, INC.

29 NEW YORK AVENUE, WESTBURY, LONG ISLAND, N. Y.
EDgewood 4-3200

**Microwave & Antenna
Instrumentation and
Systems Components**

**ENGINEERING REPRESENTATIVES
FOR:**

ANTLAB, INC.
BURMAC ELECTRONICS
CASCADE RESEARCH CO.
FENWAL ELECTRONICS, INC.
GLASS-TITE INDUSTRIES, INC.
B. F. GOODRICH SPONGE
PRODUCTS DIVISION
GORHAM MANUFACTURING
COMPANY
ITT INDUSTRIAL PRODUCTS
DIVISION
MICROWAVE ELECTRONICS
CORPORATION
MICROWAVE ELECTRONIC TUBE
CORPORATION
POLYTECHNIC RESEARCH &
DEVELOPMENT CO., INC.



501 Broadway
Hicksville, N.Y.
WElls 8-8644

1403 Arnold Ave.
Roslyn, Penna.
OLdfield 9-2938

AERO GEO ASTRO Corporation

Alexandria, Virginia

ELECTRONIC SYSTEMS AND COMPONENTS

RECONNAISSANCE

SURVEILLANCE

GUIDANCE

COMMUNICATIONS

COUNTERMEASURES

Opportunities for Mature Scientists and Engineers

DO NOT WRITE IN THIS SPACE



ACRO GEO ASTRO CORPORATION

Wilmington, Virginia

ELECTRONIC SYSTEMS AND COMPONENTS

RESEARCH AND DEVELOPMENT

TELEVISION

COMMUNICATIONS

COMMUNICATIONS

COMMUNICATIONS

OFFICE OF THE DIRECTOR OF THE ARMY AND AIR FORCE